



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TUGAS AKHIR - KI141502**

**PEMERINGKATAN PERGURUAN TINGGI MENGGUNAKAN  
METODE *PROBABILISTIC LATENT SEMANTIC ANALYSIS*  
(pLSA) UNTUK MENGUKUR TINGKAT KESIAPTERAPAN  
TEKNOLOGI PERGURUAN TINGGI DI INDONESIA**

**Donny Aliyanto**  
**NRP 5113 100 165**

**Dosen Pembimbing I**  
**Prof. Drs. Ec. Ir. Rivanarto Sarno, M.Sc.,Ph.D**

**Dosen Pembimbing II**  
**Bagus Setya Rintyarna, S.T., M.Kom.**

**JURUSAN TEKNIK INFORMATIKA**  
**Fakultas Teknologi Informasi**  
**Institut Teknologi Sepuluh Nopember**  
**Surabaya 2017**

*[Halaman ini sengaja dikosongkan]*



**TUGAS AKHIR - KI141502**

**PEMERINGKATAN PERGURUAN TINGGI MENGGUNAKAN  
METODE *PROBABILISTIC LATENT SEMANTIC ANALYSIS*  
(pLSA) UNTUK MENGUKUR TINGKAT KESIAPTERAPAN  
TEKNOLOGI PERGURUAN TINGGI DI INDONESIA**

**Donny Aliyanto  
NRP 5113 100 165**

**Dosen Pembimbing I  
Prof. Drs. Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D.**

**Dosen Pembimbing II  
Bagus Setya Rintyarna, S.T., M.Kom.**

**JURUSAN TEKNIK INFORMATIKA  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2017**

*[Halaman ini sengaja dikosongkan]*



**FINAL PROJECT - KI141502**

**UNIVERSITY RANKING USING *PROBABILISTIC LATENT SEMANTIC ANALYSIS* (pLSA) METHOD TO MEASURE APPLICATION LEVEL TECHNOLOGY OF UNIVERSITY IN INDONESIA**

**Donny Aliyanto**  
**NRP 5113 100 165**

**Supervisor I**  
**Prof. Drs. Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D.**

**Supervisor II**  
**Bagus Setya Rintyarna, S.T., M.Kom.**

**INFORMATICS DEPARTMENT**  
**Faculty of Information Technology**  
**Institut Teknologi Sepuluh Nopember**  
**Surabaya 2017**

*[Halaman ini sengaja dikosongkan]*

## LEMBAR PENGESAHAN

# PEMERINGKATAN PERGURUAN TINGGI MENGGUNAKAN METODE *PROBABILISTIC LATENT SEMANTIC ANALYSIS* (pLSA) UNTUK MENGUKUR TINGKAT KESIAPTERAPAN TEKNOLOGI PERGURUAN TINGGI DI INDONESIA

## TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada  
Rumpun Mata Kuliah Manajemen Informasi  
Program Studi S-1 Jurusan Teknik Informatika  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh  
**DONNY ALIYANTO**  
NRP. 5113 100 165

Disetujui oleh Dosen Pembimbing Tugas Akhir:

Prof. Drs. Ec. Ir. RIYANARTO SARNO  
M.Sc., Ph.D.  
NIP: 19590803 198601 1 001

BAGUS SETYA RINTYARNA  
M.Kom.  
NIK: 0509502



**SURABAYA**  
**JULI, 2017**

*[Halaman ini sengaja dikosongkan]*



# **PEMERINGKATAN PERGURUAN TINGGI MENGUNAKAN METODE *PROBABILISTIC LATENT SEMANTIC ANALYSIS* (pLSA) UNTUK MENGUKUR TINGKAT KESIAPTERAPAN TEKNOLOGI PERGURUAN TINGGI DI INDONESIA**

<b>Nama</b>	<b>: Donny Aliyanto</b>
<b>NRP</b>	<b>: 5113100165</b>
<b>Jurusan</b>	<b>: Teknik Informatika – FTIf ITS</b>
<b>Dosen Pembimbing I</b>	<b>: Prof. Drs. Ec. Ir. Riyanarto Sarno, M.Sc.,Ph.D.</b>
<b>Dosen Pembimbing II</b>	<b>: Bagus Setya Rintyarna, S.T., M.Kom.</b>

## **Abstrak**

*Pemeringkatan perguruan tinggi merupakan sebuah cara untuk mengukur kualitas dan kesiapan perguruan tinggi dalam menjalankan proses pembelajaran kepada mahasiswa yang dilihat dari berbagai macam kriteria penilaian. Pemeringkatan perguruan tinggi menjadi salah satu hal yang penting untuk mengukur reputasi perguruan tinggi di dunia internasional maupun di dalam negeri. Semakin baik reputasi, maka akan semakin baik kualitas pendidikan perguruan tinggi tersebut. Aspek kriteria penilaian bergantung pada berbagai macam penilaian yang dapat disimpulkan menjadi dua jenis yaitu kualitatif dan kuantitatif. Kriteria penilaian yang akan ditentukan harus mampu dijangkau oleh perguruan tinggi agar menghasilkan pemeringkatan yang seimbang dan ideal.*

*Dalam Tugas Akhir ini diusulkan sebuah gagasan baru sistem pemeringkatan perguruan tinggi di Indonesia menggunakan metode Probabilistic Latent Semantic Analysis (PLSA) untuk mengoptimalkan kriteria penilaian yang bersifat kualitatif, yaitu kriteria reputasi akademik dengan menggunakan metode topik modeling untuk menggali topik utama dari dokumen yang berisi*

*abstrak penelitian akademisi di setiap perguruan tinggi, yang sebelumnya dikerjakan secara manual dengan melakukan survey oleh manusia. Menggunakan 800 dokumen abstrak yang didapatkan dari Google Scholar, 9 corpus kata tingkat kesiapterapan teknologi untuk dataset kriteria yang bersifat kualitatif, serta basis data online dari Science and Techonogy Index (SINTA) dan pangkalan data PDDIKTI untuk dataset kriteria yang bersifat kuantitatif. Pengoptimalan kriteria kualitatif bertujuan untuk menghasilkan hasil pemeringkatan yang efisien dan sedernaha dari yang sebelumnya.*

*Hasil implementasi sistem pemeringkatan Perguruan Tinggi menunjukkan bahwa kinerja sistem secara keseluruhan telah mendekati Ground Truth yaitu pemeringkatan QS University Rankings dengan tingkat similarity sistem sebesar 78,33%, serta tingkat distance metode PLSA dengan Ground Truth sebesar 16 dan tingkat keberhasilan peringkat dalam toleransi sebesar 88,89%. Dengan hasil ini menunjukkan bahwa sistem telah berhasil melakukan optimasi kriteria kualitatif reputasi akademik. Diharapkan pengembangan proses pemeringkatan selanjutnya dapat mengacu pada prinsip optimasi kriteria yang bersifat kualitatif serta mengoptimalkan kembali proses topik modeling agar menjadi lebih efisien dalam menghemat waktu kerja dan biaya serta mendapatkan hasil yang akurat.*

***Kata kunci: Kualitas Perguruan Tinggi, Kualitatif, Kuantitatif, Probabilistic Latent Semantic Analysis, PLSA, Supervised PLSA, Optimasi, Pemeringkatan Perguruan Tinggi, QS University Rankings, Reputasi***

# **UNIVERSITY RANGKING USING PROBABILISTIC LATENT SEMANTIC ANALYSIS (pLSA) METHOD TO MEASURE APPLICATION LEVEL TECHNOLOGY OF UNIVERSITY IN INDONESIA**

<b>Student Name</b>	<b>: Donny Aliyanto</b>
<b>NRP</b>	<b>: 511310065</b>
<b>Major</b>	<b>: Informatics Department – FTIf ITS</b>
<b>Supervisor I</b>	<b>: Prof. Drs. Ec. Ir. Riyanarto Sarno, M.Sc., Ph.D.</b>
<b>Supervisor II</b>	<b>: Bagus Setya Rintyarna, S.T., M.Kom.</b>

## ***Abstract***

*University Ranking is a way to measure quality and readiness of college to apply their learning process to college student that seen from some criteria assessment. Universities ranking be important to measure college reputation in the outside world and within the country. With a good reputation, then can show if quality college is well. Not only that, but can become reflection of quality national education in the international outlook. Aspect of assessment relies variety of assessment that can be summarized into two types that is qualitative and quantitative. Assessment aspect specified should be able to be reached by the college in order to produce a balanced rangkings.*

*In this final project proposed a new idea for rating system universities in Indonesia using Probabilistic Latent Semantic Analysis (pLSA) method to optimize the assessment criteria are qualitative, in order to generate the rating more efficient and simple. In addition, to measure the application level technology in university. Using 800 abstract documents obtained from Google Scholar, 9 corpus words technology readiness level (TKT) for qualitative criteria datasets, as well as an online database of Science and Techonogy Index (SINTA) and PDDIKTI database for quantitative criteria datasets. Optimization of qualitative criteria*

*aims to produce more efficient and more efficient results than ever before. In addition, measuring the level of technological readiness in Higher Education to run the learning process.*

*The implementation of university ranking system shows that the result is close to the Ground Truth, which is QS University Rankings ranking with similarity value is 78,33%, gap distance method with Ground Truth is 16, and tolerance of difference rank is 88,89%. With these results indicate that the system has successfully optimized the qualitative criteria of academic reputation with some notes in the topic modeling method as an evaluation material in future development to approximate existing results. It is expected that the development of the rating process can then refer to the principle of optimization of qualitative criteria to be more efficient and simpler in saving work time.*

***Keywords: Probabilistic Latent Semantic Analysis, PLSA, Supervised PLSA, University Rankings, University Quality, QS University Rankings, Qualitative, Quantitative, Reputation, Technology Readiness Level***

## KATA PENGANTAR

Puji syukur kehadiran Allah SWT, karena atas berkat rahmat-Nya, penulis dapat menyelesaikan Tugas Akhir yang berjudul:

***“Pemeringkatan Perguruan Tinggi Menggunakan Metode Probabilistic Latent Semantic Analysis (pLSA) Untuk Mengukur Tingkat Kesiapterapan Teknologi Perguruan Tinggi Di Indonesia”***

Pengerjaan Tugas Akhir ini menjadi sebuah sarana untuk penulis memperdalam ilmu yang telah didapatkan selama menempuh pendidikan di kampus perjuangan Institut Teknologi Sepuluh Nopember Surabaya, khususnya dalam disiplin ilmu Teknik Informatika. Terselesaikannya buku Tugas Akhir ini tidak terlepas dari bantuan dan dukungan semua pihak. Pada kesempatan kali ini penulis ingin mengucapkan terima kasih kepada:

1. Bapak, Ibu, adik dan keluarga yang selalu memberikan dukungan penuh untuk menyelesaikan Tugas Akhir ini.
2. Bapak Riyanarto Sarno dan bapak Bagus Setya Rintyarna selaku dosen pembimbing yang telah bersedia meluangkan waktu untuk memberikan petunjuk selama proses pengerjaan Tugas Akhir ini.
3. Bapak dan Ibu dosen Jurusan Teknik Informatika ITS yang telah banyak memberikan ilmu dan bimbingan yang tak ternilai harganya bagi penulis.
4. Seluruh staf dan karyawan FTIf ITS yang banyak memberikan kelancaran administrasi akademik kepada penulis.
5. Segenap dosen rumpun mata kuliah Manajemen Informasi.
6. Rekan-rekan mahasiswa S2 Teknik Informatika yang telah banyak meluangkan waktu dan bersedia membantu penulis dalam memahami pengerjaan Tugas Akhir ini.

7. Rekan-rekan pengerjaan satu tim: Eko Putro Fitrianto, Anugerah Yulindra Setyaji, Arga Lancana dalam penyelesaian Tugas Akhir
8. Rekan-rekan kontrakan berprestasi dan Blogger Bangkalan : Mufarrohah, Risky Junaidi, Julio Anthony Leonard, Daniel Bintar Sijabat, Irsyad Iswanda, Nurul Mahmud, Muhamad Luthfie La Roeha, Aryandha, Romi, Daniel Fablius, Nugroho Wicaksono, Dewangga Winardi, M Fahmi Purnomo, John Stephanus Peter
9. Teman-teman seperjuangan anak didik Tugas Akhir Prof. Rryanarto Sarno yaitu Ahmad Zaenal Mustofa, Renanda Agustiantoro, Faisal Yanuar, Faisal Anugerah, Andi Putra, Hariyanto, Aldhiaz Fathra, Arfian Fidianoro, dan M. Fariz Ponighzwa.
10. Teman-teman Teknik Informatika Angkatan 2013 yang selalu mendukung, menyemangati, membantu dan mendengarkan suka duka selama proses pengerjaan Tugas Akhir.
11. Serta semua pihak yang turut membantu penulis dalam menyelesaikan Tugas Akhir ini.

Penulis menyadari masih ada kekurangan dalam penyusunan Tugas Akhir ini. Penulis mohon maaf atas kesalahan, kelalaian maupun kekurangan dalam penyusunan Tugas Akhir ini. Kritik dan saran yang membangun dapat disampaikan sebagai bahan perbaikan ke depan

Surabaya, Juli 2017

Donny Aliyanto

## DAFTAR ISI

LEMBAR PENGESAHAN.....	vii
Abstrak .....	ix
<i>Abstract</i> .....	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI .....	xv
DAFTAR GAMBAR.....	xxi
DAFTAR TABEL .....	xxiii
DAFTAR KODE SUMBER.....	xxvii
1 BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan .....	4
1.5 Manfaat.....	5
1.6 Metodologi .....	5
1.7 Sistematika Penulisan.....	7
2 BAB II DASAR TEORI.....	11
2.1 <i>University Rankings</i> .....	11
2.2 <i>QS World University Rankings</i> .....	11
2.3 Teks Mining .....	17
2.4 Teknik Grabbing .....	18
2.4.1 Fungsi cURL .....	18
2.4.2 Fungsi File_get_contents .....	19
2.4.3 Simple HTML Dom Parser .....	19

2.4.4	Regular Expression (ReGex)	20
2.5	Pusat Data <i>Scopus</i>	20
2.6	<i>Google Scholar</i>	21
2.7	<i>Science and Technology Index (SINTA)</i>	22
2.8	Pangkalan Data Pendidikan Tinggi (PDDIKTI)	23
2.9	<i>Alexa Traffict Rank</i>	24
2.10	<i>MOZ Analytics</i>	24
2.11	JAVA	24
2.12	PHP dan HTML	25
2.13	<i>Database Management System (DBMS) MySQL</i>	25
2.14	<i>Part-Of-Speech Tagger (POS Tagger)</i>	26
2.15	<i>Bloom Taxonomy</i>	28
2.16	<i>LSA (Latent Semantic Analysis)</i>	29
2.17	<i>pLSA ( Probabilistic Latent Semantic Analysis)</i>	33
2.18	Fase Teks Mining (Metode pLSA) Studi Kasus Pemeringkatan Perguruan Tinggi	37
2.19	Akreditasi Perguruan Tinggi	41
2.20	<i>Tingkat Kesiapterapan Teknologi (TKT)</i>	42
3	BAB III ANALISIS DAN PERANCANGAN	46
3.1.	Tahap Analisis	46
3.1.1	Deskripsi Umum	46
3.1.2	Aktor Dalam Sitem	47
3.1.3	Spesifikasi Kebutuhan Sistem	47
3.1.4	Kasus Penggunaan	48
3.1.5	Conceptual Data Model (CDM)	59
3.1.6	Physical Data Model (PDM)	60



3.1.7	Label Kelas dalam Teks Mining Metode PLSA .....	61
3.1.8	Ground Truth Pengujian Sistem.....	62
3.1.9	Dataset Pengujian Sistem.....	66
3.2	Tahap Perancangan .....	70
3.2.1	Perancangan Basis Data.....	70
3.2.2	Perancangan Antarmuka .....	80
3.2.3	Perancangan Kriteria Penilaian Perguruan Tinggi ....	83
3.2.4	Proses Sistem Aplikasi Pengguna (Diagram Aplikasi) .....	88
3.2.5	Proses Kerja Sistem Secara Keseluruhan (Flowchart Sistem) .....	89
3.2.6	Proses Teks Mining Metode PLSA (Flowchart Metode PLSA) .....	90
4	BAB IV IMPLEMENTASI SISTEM.....	92
4.1	Lingkungan Impelementasi .....	92
4.2	Implementasi Syntax Basis Data.....	93
4.2.1	Create Database “databse_qs_rank” .....	93
4.2.2	Create Table “universitas” .....	93
4.2.3	Create Table “reputasi_akademik” .....	93
4.2.4	Create Tabel “akreditasi” .....	94
4.2.5	Create Tabel “jumlah_mahasiswa” .....	94
4.2.6	Create Tabel “jumlah_tenaga_pengajar_jabatan_ akademik” .....	95
4.2.7	Create Tabel “jumlah_tenaga_pengajar_tetap” .....	95
4.2.8	Create Tabel “jumlah_tenaga_pengajar_tidak_tetap”	96
4.2.9	Create Tabel “jumlah_sitasi_sinta” .....	97
4.2.10	Create Tabel “jumlah_author_sinta” .....	97

4.2.11	Create Tabel “precense” .....	98
4.2.12	Create Tabel “impact” .....	98
4.2.13	Create Tabel “alexa_rank” .....	99
4.2.14	Create Tabel “kumpulan_skor_final” .....	99
4.2.15	Create Tabel “4icu” .....	100
4.2.16	Create Tabel “dikti” .....	101
4.2.17	Create Tabel “qs_general” .....	101
4.2.18	Create Tabel “qs_asia” .....	101
4.2.19	Create Tabel “qs_plsa” .....	102
4.2.20	Create Tabel “webometrics” .....	102
4.3	Tahap Prapemrosesan.....	102
4.3.1	Preprocessing.....	103
4.3.2	Membaca Dokumen .....	104
4.3.3	Membaca Kata dalam Dokumen.....	104
4.4	Tahap Perhitungan Probabilitas Topik terhadap Dokumen dan Kata .....	105
4.5	Tahap Perhitungan Random Probability .....	106
4.6	Tahap Perhitungan <i>Expectation Maximization</i> .....	107
4.7	Implementasi Antarmuka Pengguna .....	112
4.7.1	Antarmuka Home (index).....	112
4.7.2	Antarmuka Mencari Perguruan Tinggi Berdasarkan Potongan Nama .....	112
4.7.3	Antarmuka Hasil Pencarian Perguruan Tinggi Berdasarkan Keyword Potongan Nama .....	113
4.7.4	Antarmuka Melihat Detail Informasi Perguruan Tinggi .....	113

4.7.5	Antarmuka Melihat Latar Belakang Sistem Pemeringkatan .....	114
4.7.6	Antarmuka Melihat Informasi Metode Penilaian ....	115
4.7.7	Antarmuka Pencarian Peringkat Perguruan Tinggi Berdasarkan Region.....	116
4.7.8	Antarmuka Hasil Pencarian Perguruan Tinggi Berdasar Region Wilayah Perguruan Tinggi .....	117
4.7.9	Antarmuka Pembobotan Pemeringkatan Perguruan Tinggi .....	117
4.7.10	Antarmuka Hasil Pemeringkatan Perguruan Tinggi dengan Berbagai Macam Tipe Pemeringkatan .....	118
4.7.11	Antarmuka Error Handling Pembobotan Kriteria Pemeringkatan .....	121
5	BAB V PENGUJIAN DAN EVALUASI .....	122
5.1	Lingkungan Uji Coba .....	122
5.2	Data Uji Coba.....	122
5.2.1	Data Dokumen Abstrak Paper Akademisi .....	122
5.2.2	Data Corpus Taxonomy Bloom.....	125
5.2.3	Data Corpus Tingkat Kesiapterapan Teknologi (TKT) .....	126
5.2.4	Rekap Jumlah Mahasiswa dan Tenaga Pengajar.....	130
5.2.5	Data Online Science and Technologu Index (SINTA) .....	132
5.2.6	Data Online Analisa Webhost .....	134
5.3	Skenario Pengujian.....	135
5.3.1	Skenario Pengambilan Data Online Menggunakan Teknik <i>Grabbing</i> .....	135
5.3.2	Skenario Pengambilan Data pada Tabel Database ..	141

5.3.3	Skenario Pengujian Dataset Menggunakan Teks Mining Metode PLSA .....	144
5.3.4	Skenario Pengujian Sitem Berdasarkan Hasil Pembobotan Akhir Sesuai <i>Ground Truth QS Worlds University Rankings</i> .....	164
5.4	Evaluasi .....	165
6	BAB VI KESIMPULAN DAN SARAN .....	167
6.1	Kesimpulan .....	167
6.2	Saran.....	168
	DAFTAR PUSTAKA.....	169
	LAMPIRAN .....	172
	Kode Sumber.....	172
	BIODATA PENULIS.....	190

## DAFTAR GAMBAR

Gambar 2.3-1 Proses Utama Teks Mining .....	17
Gambar 2.22-1 Hirarki Tingkat Kesiapterapan Teknologi (TKT) .....	43
Gambar 1-1 Diagram Kasus Penggunaan.....	49
Gambar 1-2 Diagram Aktifitas Use Case UC-001 .....	52
Gambar 1-3 Diagram Aktifitas Use Case UC-002 .....	54
Gambar 1-4 Diagram Aktifitas Use Case UC-003 .....	55
Gambar 1-5 Diagram Aktifitas Use Case UC-004 .....	56
Gambar 1-6 Diagram Aktifitas Use Case UC-005 .....	58
Gambar 1-7 Conceptual Data Model.....	59
Gambar 1-8 Physical Data Model .....	60
Gambar 3.2-1 Alur Menentukan Reputasi Akademik .....	83
Gambar 3.2-2 Alur Menentukan Skor Akreditasi .....	84
Gambar 3.2-3 Alur Menentukan Skor Rasio Mahasiswa dan Tenaga Penagajar Aktif .....	85
Gambar 3.2-4 Alur Menentukan Skor Jumlah Sitasi Akademisi Penelitian .....	86
Gambar 3.2-5 Diagram Aplikasi .....	88
Gambar 3.2-6 Flowchart Sistem Keseluruhan.....	89
Gambar 3.2-7 Flowchart Metode PLSA.....	90
Gambar 4.4-1 Antarmuka Home .....	112
Gambar 4.4-2 Antarmuka Pencarian Nama.....	113
Gambar 4.4-3 Antarmuka Hasil Pencarian.....	113
Gambar 4.4-4 Antarmuka Detail Informasi.....	114
Gambar 4.4-5 Antarmuka Informasi Latar Belakang .....	114
Gambar 4.4-6 Antarmuka Informasi Metode Penilaian .....	115
Gambar 4.4-7 Antarmuka Metode Penilaian QS General .....	115
Gambar 4.4-8 Antarmuka Metode Penilaian QS Asia.....	116
Gambar 4.4-9 Antarmuka Metode Penilaian Webometrics.....	116
Gambar 4.4-10 Antarmuka Pencarian Berdasar Region.....	117
Gambar 4.4-11 Antarmuka Hasil Pencarian Berdasar Region ..	117
Gambar 4.4-12 Antarmuka Pembobotan Pemeringkatan .....	118
Gambar 4.4-13 Antarmuka Hasil Pemeringkatan QS General..	119
Gambar 4.4-14 Antarmuka Hasil Pemeringkatan QS Asia .....	119

Gambar 4.4-15 Antarmuka Hasil Pemeringkatan Webometrics	119
Gambar 4.4-16 Antarmuka Hasil Pemeringkatan 4ICU.....	120
Gambar 4.4-17 Antarmuka Hasil Pemeringkatan DIKTI.....	120
Gambar 4.4-18 Antarmuka Hasil Peringkat QS PLSA .....	120
Gambar 4.4-19 Antarmuka Error Handling Pembobotan.....	121
Gambar 5.3-1 Antarmuka Pembobotan Peringkat.....	164
Gambar 5.3-2 Antarmuka Tombol Cek Pemeringkatan.....	165
Gambar 5.3-3 Antarmuka Hasil Pemeringkatan.....	165

## DAFTAR TABEL

Tabel 2.15-1 Struktur Kelas kata POS Tagging .....	26
Tabel 1-1 Daftar Kebutuhan Fungsional Sistem .....	47
Tabel 1-2 Daftar Kebutuhan Fungsional Sistem .....	49
Tabel 1-3 Case Memeasukkan Presentase Bobot Kriteria Penilaian Pemeringkatan .....	50
Tabel 1-4 Use Case Melihat Peringkat Perguruan Tinggi Berdasarkan Tipe Pemeringkatan .....	53
Tabel 1-5 Use Case Mencari Perguruan Tinggi Berdasarkan Potongan Nama .....	54
Tabel 1-6 Use Case Mencari Perguruan Tinggi Berdasarkan Region Wilayah Provinsi .....	55
Tabel 1-7 Use Case Melihat Detail Informasi Utama Perguruan Tinggi .....	56
Tabel 1-8 Penjelasan Setiap Kelas Label .....	61
Tabel 1-9 Peringkat QS World University Rankings 2015-2016	62
Tabel 1-10 Peringkat Reputasi Akademik QS 2015-2016 .....	63
Tabel 1-11 Peringkat QS World University Rankings 2016-2017 .....	63
Tabel 1-12 Peringkat Reputasi Akademik QS 2016-2017 .....	64
Tabel 1-13 Peringkat QS World University Rankings 2017-2018 .....	64
Tabel 1-14 Peringkat Reputasi Akademik QS 2017-2018 .....	65
Tabel 1-15 Contoh Hasil Analisa Manual Sebagai Ground Truth Reputasi Akademik .....	66
Tabel 1-16 Daftar Rincian Jumlah Kata Dataset Abstrak Paper Perguruan Tinggi .....	67
Tabel 1-17 Daftar Rincian Jumlah Kata Corpus Taxonomy Bloom .....	68
Tabel 1-18 Perbandingan Jumlah Kata Corpus TKT .....	69
Tabel 5.2-1 Peringkat Perguruan Tinggi .....	123
Tabel 5.2-2 Daftar Reduksi Jumlah Kata Dataset .....	124
Tabel 5.2-3 Rincian Jumlah Kata Corpus Taxonomy Bloom....	125
Tabel 5.2-4 Penggabungan Referensi Corpus Taxonomy Bloom .....	126

Tabel 5.2-5 Hasil Breakdown Menjadi 9 Corpus TKT .....	127
Tabel 5.2-6 Hasil Penambahan Kata Corpus TKT Menggunakan POS Tagging .....	128
Tabel 5.2-7 Hasil Penambahan Kata Corpus TKT Menggunakan Sinonim Word .....	129
Tabel 5.2-8 Daftar Jumlah Perguruan Tinggi Berdasarkan Jenis Wilayah Administrasi.....	131
Tabel 5.3-1 Hasil Reduksi Kata Tahap Preprocessing .....	144
Tabel 5.3-2 Tingkat Duplikasi Kata .....	146
Tabel 5.3-3 Rincian Jumlah Penambahan Kata dengan POS Tagging Setiap Perguruan Tinggi.....	148
Tabel 5.3-4 Rincian Jumlah Breakdown Hasil POS Tagging pada Corpus TKT.....	149
Tabel 5.3-5 Corpus TKT Final Berdasarkan Percobaan POS Tagging.....	150
Tabel 5.3-6 Analisa Kata Hasil Akuisisi POS Tagging 1 (-1, +1) .....	150
Tabel 5.3-7 Analisa Kata Hasil Akuisisi POS Tagging 2 (-2, +2) .....	151
Tabel 5.3-8 Analisa Kata Hasil Akuisisi POS Tagging 3 (-3, +3) .....	152
Tabel 5.3-9 Analisa Kata Hasil Akuisisi POS Tagging 4 (-4, +4) .....	153
Tabel 5.3-10 Analisa Kata Hasil Akuisisi POS Tagging 8 (-8, +8) .....	153
Tabel 5.3-11 Analisa Kata Hasil Akuisisi POS Tagging 10 (-10, +10) .....	154
Tabel 5.3-12 Label Dokumen Hasil Percobaan Teknik POS Tagging Akuisisi 1 .....	155
Tabel 5.3-18 Probabilitas Label dengan Dokumen dalam Percobaan 1 .....	157
Tabel 5.3-19 Probabilitas Label dengan Dokumen dalam Percobaan 2 .....	158
Tabel 5.3-30 Perbandingan Peringkat Ground Truth QS World University Rankings 2016-2017 dengan Hasil Percobaan 1 .....	160



Tabel 5.3-35 Perbandingan Peringkat Ground Truth QS World University Rankings 2016-2017 dengan Hasil Percobaan 2 .....	161
Tabel 5.3-37 Hasil Analisa Semua Percobaan dengan Ground Truth QS World University Rankings 2016-2017.....	162
Tabel 5.3-38 Hasil Analisa Semua Percobaan dengan Ground Truth QS World University Rankings 2017-2018.....	163
Tabel 5.4-1 Rangkuman Hasil Pengujian.....	165

*[Halaman ini sengaja dikosongkan]*

## DAFTAR KODE SUMBER

Kode Sumber 2.4-1 Syntax cURL dengan setting option url .....	19
Kode Sumber 2.4-2 Syntax Fungsi file_get_contents .....	19
Kode Sumber 2.4-3 Syntax Penggunaan Library Simple HTML Dom Parser .....	20
Kode Sumber 4.2-1 Kode Sumber Create Database.....	93
Kode Sumber 4.2-2 Kode Sumber Create Table universitas .....	93
Kode Sumber 4.2-3 Create Tabel reputasi_akademik .....	93
Kode Sumber 4.2-4 Create Tabel akreditasi.....	94
Kode Sumber 4.2-5 Create Table jumlah_mahasiswa.....	94
Kode Sumber 4.2-6 Create Table jumlah_tenaga_pengajar_jabatan_akademik .....	95
Kode Sumber 4.2-7 Create Table jumlah_tenaga_pengajar_tetap .....	95
Kode Sumber 4.2-8 Create Table jumlah_tenaga_pengajar_tidak_tetap .....	96
Kode Sumber 4.2-9 Create Table jumlah_sitasi_sinta .....	97
Kode Sumber 4.2-10 Create Table jumlah_author_sinta .....	98
Kode Sumber 4.2-11 Create Table precense .....	98
Kode Sumber 4.2-12 Create Table impact .....	99
Kode Sumber 4.2-13 Create Tabel alexa_rank.....	99
Kode Sumber 4.2-14 Create Table kumpulan_skor_final .....	99
Kode Sumber 4.2-15 Create Table kumpulan_skor_final .....	100
Kode Sumber 4.2-16 Create Table 4icu .....	101
Kode Sumber 4.2-17 Create Table dikti .....	101
Kode Sumber 4.2-18 Create Table qs_general.....	101
Kode Sumber 4.2-19 Create Table qs_asia .....	101
Kode Sumber 4.2-20 Create Table qs_plsa .....	102
Kode Sumber 4.2-21 Create Table webometrics .....	102
Kode Sumber 4.3-1 Kode Sumber Preprocessing.....	104



# **BAB I**

## **PENDAHULUAN**

Pada bab ini dibahas hal-hal yang mendasari dalam Tugas Akhir. Bahasan meliputi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, metodologi, dan sistematika penulisan Tugas Akhir.

### **1.1 Latar Belakang**

Perguruan tinggi merupakan sebuah institusi pendidikan penyelenggara pendidikan tinggi. Pendidikan tinggi menjadi hal penting karena mencerminkan kondisi intelektual masyarakat dari sebuah negara. Membutuhkan perhatian dan evaluasi lebih dari pemerintah untuk dapat menghasilkan sumberdaya manusia yang baik dari perguruan tinggi. Hal ini tentunya untuk menghasilkan generasi yang mempunyai intelektual dan keterampilan tinggi ketika bersaing di dunia kerja kelak [1]. Kualitas pendidikan tinggi yang baik menjadi prioritas dan pusat perhatian pendidikan internasional. Pencapaian tujuan strategis perguruan tinggi bergantung pada faktor kontekstual, yang meliputi kepemimpinan suatu negara, ideologi negara, finansial, sosial budaya, komunikasi dan bahasa, serta proses pengambilan keputusan dari pemerintahan. Selain itu perguruan tinggi juga didorong oleh norma yang terkait dengan lingkungan dan nilai pendidikan itu sendiri [1].

Selain itu perguruan tinggi menjadi salah satu sarana yang efektif untuk mengembangkan ilmu pengetahuan dan teknologi dengan berbagai cara, salah satu cara dengan melakukan penelitian. Penelitian merupakan salah satu wadah kreatifitas dan pengembangan potensi akademik bagi mahasiswa. Selain itu penelitian diselenggarakan untuk membuat seorang akademisi dapat menghasilkan sebuah karya dalam dunia pendidikan [1]. Jumlah populasi masyarakat suatu negara yang banyak maka harus berbanding lurus dengan pengembangan intelektual dan keterampilan generasi muda, sebagai investasi negara dalam dunia pendidikan.

Salah satu upaya untuk meningkatkan kualitas perguruan tinggi adalah dengan melihat hasil pemeringkatan perguruan tinggi di seluruh dunia [2]. Perguruan tinggi yang masuk dalam peringkat dunia akan menjadi acuan dan fokus bahwa kualitas pendidikan di negara tersebut telah bagus, dari berbagai segi kriteria. Ada banyak media publikasi pemeringkatan perguruan tinggi, salah satunya adalah publikasi tahunan *QS World University Rankings* [3]. Dalam pemeringkatan perguruan tinggi, tentunya ada banyak kriteria penilaian yang menjadikan perguruan tinggi tersebut berkualitas dan layak untuk menempati peringkat dalam daftar.

Tingkat kualitas pendidikan di Indonesia sejauh ini dapat dikatakan kurang baik. Indonesia masih berfokus pada jumlah partisipasi pendidikan, tidak pada pembangunan kualitas pendidikan. Berdasarkan posisi peringkat publikasi tahunan *QS World University Rankings*, perguruan tinggi di Indonesia masih berada pada urutan 67 regional Asia yang ditempati oleh Universitas Indonesia (UI). Berbeda jauh jika dibandingkan dengan negara tetangga, yaitu Singapura yang mempunyai jumlah penduduk lebih sedikit dari Indonesia berhasil berada pada peringkat 1 region Asia yang ditempati oleh National University of Singapore (NUS) [4]. Di Indonesia sendiri kualitas pendidikan yang menonjol hanya berada pada perguruan tinggi tertentu saja, hal ini yang menyebabkan kualitas pendidikan di Indonesia kurang merata.

Salah satu cara untuk mengetahui merata atau tidaknya kualitas pendidikan tinggi di Indonesia dengan melihat Tingkat Kesiapterapan Teknologi (TKT) dari perguruan tinggi. TKT menggambarkan tingkat kematangan perguruan tinggi dalam hal penelitian dan pengembangan teknologi [5]. Tujuan utama dari penelitian secara umum sejalan dengan tujuan perguruan tinggi itu sendiri, sebagai sarana yang efektif untuk mengembangkan ilmu pengetahuan dan pengembangan potensi sumberdaya manusia.

Kembali kepada publikasi tahunan *QS World University Rankings* terdapat berbagai macam kriteria penilaian yang secara umum menggambarkan kondisi perguruan tinggi [3], yang mana

dapat menjadi *value* untuk meningkatkan tingkat kesiapterapan (TKT) dari perguruan tinggi. Kriteria kualitatif dalam *QS World University Rankings* menjadi *value* yang paling berpengaruh dalam melihat kualitas perguruan tinggi. Kriteria ini didapatkan dengan proses manual yang teliti dan akurat terhadap hasil, sehingga membutuhkan waktu yang lama untuk menghasilkan pemeringkatan yang ideal [3]. Salah satu pilihan untuk mengoptimasi kriteria kualitatif adalah dengan menerapkan metode kecerdasan komputasional berupa metode teks mining *Probabilistic Latent Semantic Analysis* (pLSA) [6].

## 1.2 Rumusan Masalah

Rumusan masalah yang diangkat dalam Tugas Akhir ini dapat dipaparkan sebagai berikut:

1. Bagaimana mengumpulkan data masukan sebagai bahan penilaian kriteria perguruan tinggi?
2. Bagaimana menghasilkan skor akhir dari kriteria penilaian perguruan tinggi yang bersifat kualitatif menggunakan topic modelling metode *Probabilistic Latent Semantic Analysis* (pLSA)?
3. Bagaimana menghitung skor akhir dari data kriteria penilaian perguruan tinggi yang bersifat kuantitatif?
4. Bagaimana menentukan skor akhir secara keseluruhan untuk sebuah perguruan tinggi?
5. Bagaimana menganalisa hasil keluaran sistem dengan *Ground Truth QS University Rankings* yang sudah ditetapkan sehingga menghasilkan hasil akurasi?

## 1.3 Batasan Masalah

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki beberapa batasan sebagai berikut:

1. Bahasa pemrograman yang digunakan adalah Python untuk proses *teks mining* dokumen abstrak *paper*, PHP dan HTML untuk membangun sistem *grabbing* dan membangun desain antarmuka.

2. Database management sistem (DBMS) yang digunakan adalah MySQL database.
3. Perguruan tinggi yang dinilai hanya perguruan tinggi dalam wilayah negara Indonesia.
4. Menggunakan acuan kriteria dari media publikasi tahunan *QS World University Rankings*.
5. Data sebagai dasar daftar perguruan tinggi, data total mahasiswa, data total tenaga kerja perguruan tinggi berdasarkan data pada PDDIKTI.
6. Data total publikasi akademisi sebagai salah satu kriteria penilaian kuantitatif berdasarkan database *Science and Technology Index* (SINTA) yang mengadopsi database dari Scopus dan Google Scholar.
7. Data akreditasi perguruan tinggi berdasarkan data dari Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT).
8. Metode *Probabilistic Latent Semantic Analysis* (pLSA) untuk mengidentifikasi topik dari dokumen abstrak *paper* akademisi perguruan tinggi yang akan menjadi skor nilai reputasi akademik yang bersifat kualitatif.
9. Menggunakan 50 dataset berupa dokumen abstrak *paper* akademisi dari masing-masing perguruan tinggi dengan kategori rumpun teknik.
10. Menggunakan corpus *Taxonomy Bloom* sebagai dasar penetapan corpus kata (*term*) untuk setiap level Tingkat Kesiapterapan Teknologi (TKT).

## 1.4 Tujuan

Tujuan dari pembuatan Tugas Akhir ini sebagai berikut:

1. Mampu menerapkan teknik *grabbing* terhadap data online berupa informasi *numeric* sebagai masukan kriteria penilaian perguruan tinggi yang bersifat kuantitatif.
2. Mampu menerapkan metode *Probabilistic Latent Semantic Analysis* (pLSA) untuk memodelkan topik dari sebuah dokumen abstrak *paper* akademisi perguruan tinggi sebagai masukan kretria penilaian yang bersifat kualitatif.



3. Mampu menghasilkan skor dari setiap kriteria penilaian perguruan tinggi, serta skor final penilaian secara keseluruhan/kumulatif dari setiap kriteria.
4. Mampu menganalisa hasil keluaran sistem pemeringkatan dengan *Ground Truth QS World University Rankings*.
5. Mampu menyajikan informasi hasil pemeringkatan perguruan tinggi dalam antarmuka yang jelas dan mudah untuk dimengerti dan diakses oleh pengguna perangkat lunak website (*End User*).

## 1.5 Manfaat

Tugas Akhir ini diharapkan dapat membantu menghasilkan pemeringkatan perguruan tinggi di Indonesia yang mudah, efektif, dan sederhana berdasarkan Tingkat Kesiapterapan Teknologi (TKT) yang menjadi dasar pemetaan tingkat kematangan teknologi perguruan tinggi di Indonesia. Mengoptimasi kriteria diharapkan mampu menekan waktu kerja (*running time*) dan biaya (*cost*) untuk proses pemeringkatan perguruan tinggi.

Penilaian kriteria yang dibahas mengacu pada publikasi pemeringkatan perguruan tinggi internasional tahunan *QS World University Rankings* yang dapat digunakan sebagai tolok ukur kualitas pendidikan tinggi di Indonesia.

## 1.6 Metodologi

Tahapan-tahapan yang ditempuh dalam pengerjaan Tugas Akhir ini sebagai berikut:

### 1. Studi literatur

Pada tahap ini dilakukan studi literatur dengan mencari referensi yang berfokus pada pembahasan teknik *grabbing* data online dari website yang berfungsi sebagai penyedia informasi kriteria penilaian yang bersifat kuantitatif, mencari referensi proses teks mining metode *Probabilistic Latent Semantic Analysis (pLSA)* untuk proses topik modelling dokumen abstrak *paper* akademisi perguruan tinggi yang menjadi salah satu kriteria penilaian

yang mempunyai bobot ketergantungan paling tinggi, serta mencari referensi literatur lain yang membahas mengenai dasar-dasar dan proses pemeringkatan perguruan tinggi yang sudah ada sebelumnya.

## **2. Analisis dan Perancangan Sistem**

Pada tahap ini dijabarkan mengenai proses diagram kasus sistem (*use case*) dan aktor yang terlibat dalam alur sistem. Aktor yang menjadi pelaku dalam sistem adalah pengguna perangkat lunak website (*end user*). Kemudian beberapa kebutuhan fungsional dari sistem ini antara lain:

- i. Memasukkan prosentase bobot kriteria penilaian pemeringkatan.
- ii. Melihat peringkat perguruan tinggi berdasarkan tipe pemeringkatan.
- iii. Mencari perguruan tinggi berdasarkan potongan nama perguruan tinggi.
- iv. Mencari perguruan tinggi berdasarkan region wilayah provinsi.
- v. Melihat detail informasi akademik perguruan tinggi.

## **3. Implementasi Sistem**

Pada tahap ini dilakukan pembuatan perangkat lunak untuk sistem server (*developer*) berupa penerapan teknik *grabbing* pada pengambilan data online, proses teks mining menggunakan metode *Probabilistic Latent Semantic Analysis* (pLSA) untuk mengganti kriteria reputasi akademik yang sebelumnya dikerjakan secara manual, serta membangun database dan antarmuka untuk menampilkan informasi akhir kepada pengguna.

## **4. Uji Coba dan Evaluasi**

Pada tahap ini dilakukan pengujian terhadap perangkat lunak dengan menggunakan dataset final berupa

50 dokumen abstrak *paper* akademisi dari masing-masing perguruan tinggi sebagai pengujian kriteria penilaian yang sifat kualitatif, kemudian menggunakan data online dari database akademisi *Science and Technology Index* (SINTA) sebagai pengujian kriteria penilaian yang bersifat kuantitatif yang mempunyai nilai berupa numeric, serta menggunakan data hasil analisa Backlink, Domain Authority dan Page Authority dari kriteria berbasis analisa *webhost*.

Uji coba dan evaluasi dilakukan untuk mengevaluasi jalannya perangkat lunak, mengevaluasi fitur utama, mengevaluasi fitur tambahan, mencari kesalahan antara hasil peringkat yang dihasilkan oleh sistem perangkat lunak dengan peringkat *Ground Truth* yang sudah ditetapkan, menganalisa faktor-faktor yang mempengaruhi kebenaran sistem, dan mengadakan perbaikan jika ada kekurangan.

## **5. Penyusunan Buku Tugas Akhir**

Pada tahap ini dilakukan pendokumentasian dan pelaporan dari seluruh konsep, dasar teori, implementasi sistem, rancangan antarmuka, proses lain yang telah dilakukan, hasil-hasil dan analisa yang telah didapatkan selama pengerjaan Tugas Akhir.

### **1.7 Sistematika Penulisan**

Buku Tugas Akhir ini bertujuan untuk mendapatkan informasi dan gambaran dari pengerjaan Tugas Akhir ini. Selain itu, diharapkan dapat berguna untuk pembaca yang tertarik untuk melakukan pengembangan lebih lanjut. Secara garis besar, buku Tugas Akhir terdiri atas beberapa bagian seperti berikut:

**Bab I Pendahuluan**

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan, manfaat, metodologi yang digunakan, dan sistematika penyusunan Tugas Akhir.

**Bab II Dasar Teori**

Bab ini membahas teori sebagai penunjang yang berhubungan dengan pokok pembahasan yang menjadi dasar dari pembuatan Tugas Akhir ini.

**Bab III Analisis dan Perancangan Sistem**

Bab ini membahas mengenai perancangan perangkat lunak. Perancangan perangkat lunak meliputi perancangan alur, proses dan perancangan antarmuka pada perangkat lunak.

**Bab IV Implementasi**

Bab ini membahas mengenai implementasi dari perancangan perangkat lunak dan implementasi fitur-fitur penunjang perangkat lunak.

**Bab V Uji Coba dan Evaluasi**

Bab ini membahas pengujian dengan metode pengujian subjektif untuk mengetahui penilaian aspek kegunaan (*usability*) dari perangkat lunak, serta melakukan pengujian fungsionalitas yang dibuat dengan memperhatikan keluaran yang dihasilkan, serta evaluasi terhadap fitur-fitur perangkat lunak.

**Bab VI Kesimpulan**

Bab ini berisi kesimpulan dari hasil pengujian yang dilakukan. Bab ini membahas saran-saran untuk pengembangan perangkat lunak lebih lanjut.

**Daftar Pustaka**

Merupakan daftar referensi yang digunakan dalam proses membangun dan mengembangkan perangkat lunak dalam Tugas Akhir.

*[Halaman ini sengaja dikosongkan]*

## **BAB II**

### **DASAR TEORI**

Pada bab ini diuraikan mengenai dasar-dasar teori yang digunakan dalam pengerjaan Tugas Akhir dengan tujuan untuk memberikan gambaran secara umum terhadap penelitian yang dikerjakan.

#### **2.1 *University Rankings***

Peringkat perguruan tinggi merupakan salah satu cara untuk mengukur kualitas pendidikan di sebuah negara yang diterapkan oleh seluruh perguruan tinggi di dunia. Namun saat ini tujuan pemeringkatan perguruan tinggi cenderung berbeda dari awalnya yaitu menjadi salah satu cara untuk melihat kepopuleran perguruan tinggi tersebut di mata dunia internasional. Dimana perguruan tinggi yang mempunyai posisi peringkat tinggi akan menjadi perhatian bahwa pendidikan tinggi di negara tersebut berkualitas [1]- [2].

Hingga saat ini, ada berbagai macam publikasi pemeringkatan perguruan tinggi yang populer untuk memberikan hasil pemeringkatan perguruan tinggi di seluruh dunia [7]- [8], diantaranya:

1. QS World University Rankings.
2. Timer Higher Education World University Rankings (THE).
3. Academics Ranking of World University (ARWU).
4. Global Universities Rangking by US News.
5. Reuters World's Top 100 Innovative University.
6. SCImago Institution Rankings.
7. Webometrics.
8. 4ICU (4 International Colleges and Universities).

#### **2.2 *QS World University Rankings***

*Quacquarelli Symonds* atau lebih sering disebut dengan *QS World University Rangking* merupakan publikasi peringkat perguruan tinggi seluruh dunia yang berlokasi di Inggris [4]. *QS World University Rangking* diterbitkan setiap tahun dengan menghasilkan perubahan peringkat yang sangat sensitif, sehingga persaingan antar perguruan tinggi begitu kuat [1].

Awal mula terbentuknya *QS World University Rangking* pada tahun 2004 merupakan hasil kolaborasi dengan THE (*Times Higher Education*) dengan nama publik THE-QS, namun sejak tahun 2009 QS dan THE sepakat untuk berjalan sendiri-sendiri. QS berjalan hingga

sekarang menggunakan metode yang diterapkan sejak awal dan lebih populer disebut publikasi *QS World University Ranking*. Sedangkan THE membangun publikasi peringkat perguruan tinggi sendiri dengan nama THES yang menggunakan metode lain yang hampir serupa dengan *QS World University Ranking* [3].

*QS World University Ranking* mempunyai 5 cakupan region sebagai observasi pemeringkatan, diantaranya Asia, Amerika Latin, negara Eropa berkembang - Asia Tengah, region Arab dan BRICS (Brasil, Rusia, India, China, South Africa). Untuk menghasilkan peringkat perguruan tinggi, *QS World University Ranking* mempunyai 6 kriteria utama [4], diantaranya:

**1. Reputasi Akademik (40%)**

Kriteria ini merupakan hasil survey responden akademisi perguruan tinggi di seluruh dunia. Survey ini diharapkan akan mendapatkan hasil yang objektif dari semua responden yang tersebar diseluruh dunia. Mereka mendapatkan bagian survey sesuai dengan latar belakang akademisi tersebut. Terdapat 5 garis besar bidang latar belakang akademisi sebagai responden, yaitu bidang ilmu seni dan sastra, ilmu teknik dan teknologi, ilmu biologi dan kesehatan, ilmu alam, dan ilmu sosial [3].

**2. Reputasi Employer (10%)**

Kriteria ini merupakan hasil survey responden perusahaan yang menaungi alumni perguruan tinggi di seluruh dunia [3]. Sama dengan kriteria reputasi akademik, kriteria ini ditujukan agar memperoleh hasil yang objektif terhadap kualitas lulusan (fresh graduate) perguruan tinggi. Salah satu faktor penting agar perguruan tinggi dapat diperhatikan dalam kriteria ini adalah dengan menjalin banyak kerjasama dengan perusahaan sebagai rekan kerja maupun penelitian [2].

**3. Jumlah Mahasiswa Aktif (20%)**

Kriteria ini merupakan hasil jumlah mahasiswa yang sedang menempuh pendidikan di perguruan tinggi. Tujuan dari kriteria ini adalah untuk menganalisa pengawasan jumlah infrastruktur dengan jumlah mahasiswa, agar jumlah infrastruktur sebanding dengan jumlah mahasiswa. Perguruan tinggi yang baik merupakan perguruan tinggi yang memberikan fasilitas dan kenyamanan belajar [3].



#### **4. Jumlah Sitasi Akademisi (20%)**

Kriteria ini merupakan hasil jumlah sitasi maupun publikasi dari akademisi perguruan tinggi yang ter-*index* oleh publikasi Scopus. Sitasi/publikasi yang dihitung berusia maksimal terbit 5 tahun sebelumnya. Tujuan dari kriteria ini adalah untuk melihat seberapa banyak dampak penelitian akademisi perguruan tinggi terhadap lingkungan [3]. Karena tujuan perguruan tinggi adalah yang dapat memberikan dampak baru bagi masyarakat.

#### **5. Jumlah Tenaga Pengajar Internasional (5%)**

Kriteria ini merupakan hasil dari jumlah tenaga pengajar (staff/dosen) asing di sebuah perguruan tinggi. Lama waktu minimal tenaga pengajar asing menetap selama 3 bulan per tahun [2]. Tujuan dari kriteria ini adalah melihat kerjasama dan ketertarikan tenaga asing untuk perguruan tinggi tersebut.

#### **6. Jumlah Mahasiswa Asing (5%)**

Kriteria ini merupakan hasil dari jumlah mahasiswa asing yang belajar di perguruan tinggi, baik dalam rangka pertukaran pelajar, maupun program pendidikan lainnya [3]. Tujuan dari kriteria ini adalah untuk melihat ketertarikan mahasiswa asing terhadap metode pembelajaran perguruan tinggi [2].

Dalam penilaian kriteria *QS World University Ranking* terdapat 2 jenis kriteria, yaitu kriteria yang bersifat kualitatif dan kuantitatif. Untuk kriteria yang bersifat kualitatif merupakan kriteria yang membutuhkan pengolahan data untuk mendapatkan nilai serta cenderung bersifat deskriptif dan membutuhkan analisa terlebih dahulu. Kriteria yang termasuk dalam jenis kualitatif adalah kriteria reputasi akademik dan reputasi employer.

Sedangkan untuk kriteria yang bersifat kuantitatif merupakan kriteria yang dapat ditentukan nilainya secara langsung tanpa melakukan analisa terlebih dahulu dan tidak bersifat deskriptif. Kriteria yang termasuk dalam jenis kuantitatif adalah kriteria jumlah mahasiswa aktif, jumlah sitasi/publikasi akademisi, jumlah staff internasional dan jumlah mahasiswa asing.

Untuk kriteria pertama yang bersifat kualitatif, yaitu reputasi akademik dalam *QS World University Rankings* ada beberapa poin yang menjadi acuan untuk digali dari responden survey untuk menghasilkan penilaian yang objektif [9], diantaranya:

### 1. Biodata personal responden akademisi

Poin penilaian ini, responden diminta menjelaskan mengenai biodata personal responden yang berkaitan dengan dimana akademisi bekerja, diantaranya seperti nama lengkap, alamat email institusional, jobdesk pekerjaan, lama bekerja dalam institusi, nama institusi dimana akademisi bekerja, pendapat mengenai negara yang paling berpengaruh dalam pendidikan tinggi dan penelitian, serta menanyakan mengenai pengetahuan akademisi terhadap Massive Open Online Courses (MOOCs) [9].

### 2. Knowledge specification

Poin penilaian ini, responden diminta menjelaskan berdasar pengetahuan responden mengenai pengetahuan dasar tentang region dan bidang ilmu yang akan dibahas/ditanyakan, diantaranya seperti pengetahuan terhadap kualitas akademik berdasarkan region (amerika, asia-australia-new zealand, eropa-timur tengah-afrika atau semua region), pengetahuan terhadap kategori bidang ilmu pengajaran, pengetahuan terhadap sub bidang ilmu yang lebih spesifik, serta mengenai pembagian waktu akademisi untuk melakukan riset (*research*), pengajaran (*teaching*), administrasi (*administration*) dan lainnya dalam bentuk persen [9].

### 3. Penelitian terbaik

Poin penilaian ini, responden diminta memilih beberapa perguruan tinggi yang dianggap mempunyai tingkat penelitian (*research*) yang baik, diantaranya seperti memilih 10 institusi dalam negeri sesuai dengan region negara responden yang mempunyai kualitas penelitian terbaik, memilih 30 institusi luar negeri yang mempunyai kualitas penelitian terbaik, serta dalam memilih institusi dalam negeri maupun luar negeri tersebut harus mencakup kategori bidang ilmu yang diteliti. Ada 5 bidang ilmu, maka total responden harus memilih 50 perguruan tinggi dalam negeri terbaik yang mencakup 5 bidang ilmu, serta 150 perguruan luar negeri yang mencakup 5 bidang ilmu.

### 4. Inovasi dalam penelitian

Poin penilaian ini, responden diminta mengidentifikasi perguruan tinggi mana yang mempunyai tingkat inovasi yang baik dalam

penelitian, diantaranya seperti mengidentifikasi institusi yang paling cepat dan dinamis dalam perkembangan dan inovasi penelitian, serta bidang ilmu apa yang dimiliki institusi tersebut yang paling cepat berkembang dalam hal inovasi [9].

#### **5. Leading company**

Poin penilaian ini, responden diminta mengidentifikasi perusahaan mana yang paling berkontribusi terhadap penelitian, diantaranya seperti mengidentifikasi 5 perusahaan yang menurut responden akademisi paling berpengaruh terhadap pendidikan sesuai dengan bidang pendidikan responden, serta menjelaskan alasan memilih 5 perusahaan tersebut [9].

#### **6. MOOCs**

Poin penilaian ini, responden diminta menjelaskan mengenai penerapan MOOCs dalam institusi responden [10], diantaranya seperti apakah institusi pendidikan responden sudah menerapkan pembelajaran berbasis MOOCs, apakah institusi pendidikan responden dapat merencanakan pembelajaran berbasis MOOCs untuk 3 tahun kedepan, apakah responden telah menerapkan pembelajaran MOOCs berdasarkan inisiatif pribadi, serta pendapat responden mengenai keuntungan pembelajaran berbasis MOOCs [9].

#### **7. Profil penelitian responden akademisi**

Poin penilaian ini, responden diminta menjelaskan profil publikasi/penelitian responden selama ini jika ada, diantaranya seperti artikel jurnal, konferensi, paper/buku, kemampuan (*performance*) responden, pengalaman (*exhibitions*) responden, serta paten yang dimiliki [9].

Untuk kriteria kedua yang bersifat kualitatif yaitu reputasi employer, ada beberapa poin yang menjadi acuan untuk digali dari responden perusahaan untuk menghasilkan penilaian yang objektif [11], diantaranya:

#### **1. Biodata personal responden perusahaan**

Poin penilaian ini, responden perusahaan diminta menjelaskan mengenai biodata personal yang berkaitan dengan institusi dimana responden bekerja, diantaranya nama lengkap, nama perusahaan, alamat email resmi berdomain perusahaan, website perusahaan, jabatan pekerjaan, negara lokasi perusahaan berjalan, sektor utama perusahaan, ukuran perusahaan (kuantitas staff), kategori lulusan

yang diterima sebagai staff (internship, sarjana-magister (S1/S2), MBA), dan yang terakhir lokasi negara yang paling banyak kandidat diterima sebagai staff perusahaan [11].

## **2. Analisa perguruan tinggi bisnis terbaik**

Poin penilaian ini, responden perusahaan diminta menjelaskan mengenai pendapat mereka terhadap perguruan tinggi yang mempunyai lulusan terbaik, baik dalam hal bisnis maupun keseluruhan, diantaranya adalah memilih 10 perguruan tinggi domestik sesuai tempat tinggal responden yang mempunyai lulusan terbaik, memilih 30 perguruan tinggi luar negeri yang mempunyai lulusan terbaik, memberi tahu daftar perguruan tinggi dari staff perusahaan yang bekerja dalam perusahaan tersebut, serta bidang disiplin ilmu yang paling banyak direkrut oleh perusahaan tersebut [11].

Selain itu mengidentifikasi juga sekolah bisnis dimana responden perusahaan diminta untuk menjelaskan beberapa hal, diantaranya adalah mengidentifikasi 10 sekolah bisnis yang alumninya berpotensi direkrut oleh perusahaan, perekrutan staff bergelar MBA (Master of Business Administration), membiayai staff pada program executive MBA untuk mengembangkan potensi, menjelaskan pentingnya spesialisasi MBA dalam perusahaan, serta memilih 5 perguruan tinggi yang unggul dalam gelar MBA (ada 10 kategori spesialisasi, yang berarti ada 50 perguruan tinggi yang dipilih).

## **3. Kualifikasi untuk staff bergelar MBA**

Poin penilaian ini, responden perusahaan diminta menjelaskan pendapatnya mengenai kualifikasi staff yang bergelar MBA, diantaranya adalah memilih region negara yang mempunyai sekolah bisnis terbaik dan menciptakan dampak tinggi untuk bisnis, lalu menjelaskan mengenai ketentuan yang ideal untuk staff bergelar MBA untuk bekerja (lama waktu kerja per hari, gaji, bonus, tunjangan), divisi/posisi kerja yang sesuai dengan staff bergelar MBA, rata-rata gaji yang ditawarkan oleh perusahaan untuk freshgraduate.

## **4. Pembahasan umum**

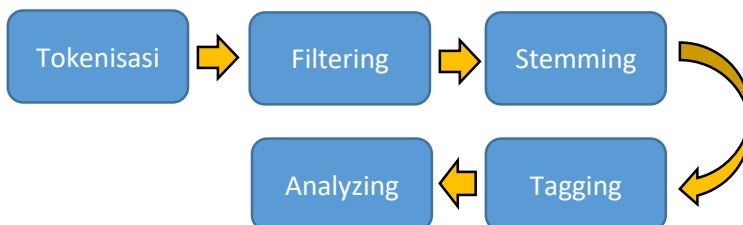
Poin penilaian ini, responden perusahaan diminta menjelaskan pendapatnya mengenai kebutuhan skill apa yang harus dimiliki

oleh calon staff (bergelar MBA maupun reguler), diantaranya adalah skill yang dibutuhkan untuk freshgraduate, seperti pengalaman internasional, pengetahuan tentang subjek, kepemimpinan, pengalaman magang, kemampuan individu, dan kemampuan bahasa. Selain itu diminta menjelaskan mengenai skill yang dibutuhkan oleh calon staff MBA, seperti kemampuan individu, interaksi dengan lingkungan, kepemimpinan, prestasi akademik, pemikiran strategis, kemampuan bahasa, kewiraswastaan, keterampilan IT, manajemen resiko, pengalaman, pemasaran, keterampilan keuangan, komunikasi, tanggung jawab sosial, dan e-business [11].

Untuk kriteria ketiga hingga keenam, poin yang dijadikan penilaian adalah jumlah dari masing-masing. Kriteria ini tidak mempunyai ketentuan poin khusus seperti kriteria sebelumnya yang bersifat kualitatif.

### 2.3 Teks Mining

Teks mining merupakan proses menggali data teks yang didapatkan dari sumber data berupa dokumen (word, pdf, kutipan, dan sejenisnya). Teks mining disebut juga dengan teknik mengekstraksi pola dari sebuah dokumen. Tujuan utama dari teks mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen dengan melakukan pencirian kata-kata yang dapat mewakili isi dari sebuah dokumen yang kemudian dapat dianalisa. Kegunaan dari penggunaan teks mining adalah sebagai pengkategorian teks (*text categorization*) dan pengelompokan teks (*text clustering*) [12]. Proses umum dalam teks mining seperti pada Gambar 2.3-1.



**Gambar 2.3-1 Proses Utama Teks Mining**

## 2.4 Teknik Grabbing

Merupakan salah satu cara untuk mengambil konten teks atau object yang ditampilkan dalam situs website yang kemudian akan diletakkan dalam website milik pribadi. Secara umum teknik ini akan menduplikasi konten website tujuan untuk ditampilkan dalam website kita, akan tetapi dari sini banyak informasi dan kegunaan lain yang dapat dikembangkan. Salah satunya adalah sebagai salah satu bahan teks mining dimana dengan tujuan untuk menganalisa konten teks dalam website tersebut. Teknik grabbing lebih banyak dikembangkan menggunakan bahasa pemrograman PHP, terdapat library yang mendukung analisa ini. Teknik grabbing dapat dikembangkan sendiri oleh developer sesuai dengan kebutuhan yaitu kegunaan dan konten situs website sasaran, oleh karena itu teknik ini bisa dikatakan tidak fleksibel karena jika ingin melakukan grabbing pada situs website tertentu maka harus merubah pola konten yang akan diambil. Beberapa cara menggunakan teknik grabbing diantaranya adalah fungsi cURL, fungsi `file_get_contents`, library simple HTML dom parser, dan pemanfaatan ReGex (*Regular Expression*).

### 2.4.1 Fungsi cURL

Merupakan sebuah fungsi yang berguna untuk mengambil konten dari sebuah situs website. cURL dapat digunakan untuk melakukan simulasi request method POST ke server dan juga menggunakan COOKIE. Fungsi ini dijalankan pada Bahasa pemrograman PHP. Keuntungan menggunakan cURL adalah beberapa situs website yang melindungi kontennya dengan menggunakan cookie dan post method, dengan menggunakan cURL dapat mengakses konten yang terproteksi tersebut. Beberapa fungsi yang diakomodasi oleh cURL:

1. `curl_init()`, untuk inisialisasi curl library.
2. `curl_setopt()`, untuk mengatur pilihan dari curl.
3. `curl_exec()`, untuk mengeksekusi/menjalankan query.
4. `curl_close()`, untuk menghentikan curl system.

```

$options = array(
    CURLOPT_CUSTOMREQUEST =>"GET",
    CURLOPT_POST           =>false,
    CURLOPT_FOLLOWLOCATION => true,
    CURLOPT_CONNECTTIMEOUT => 120,
    CURLOPT_TIMEOUT        => 120
);

$ch = curl_init( $url );
curl_setopt_array( $ch, $options );
$content = curl_exec( $ch );
curl_close( $ch );
$header['content'] = $content;

```

**Kode Sumber 2.4-1 Syntax cURL dengan setting option url**

## 2.4.2 Fungsi File\_get\_contents

Merupakan cara untuk membaca isi file ke dalam string. Karena akan menggunakan teknik pemetaan memori, jika hal ini didukung oleh server, untuk meningkatkan kinerja. Isi file dalam konteks grabbing adalah, bagian konten situs website yang akan diterjemahkan menjadi string, kemudian akan dianalisa sesuai kebutuhan dengan mengambil informasi dari tag konten bagian tertentu menggunakan ReGex (*Regular Expression*).

```
file_get_contents(path,include_path,context,start,max_length)
```

**Kode Sumber 2.4-2 Syntax Fungsi file\_get\_contents**

## 2.4.3 Simple HTML Dom Parser

Merupakan sebuah syntax dalam Bahasa pemrograman PHP yang berfungsi untuk menguraikan elemen-elemen HTML untuk diambil dan dirubah atau dihapus pada tag/elemen-elemen tertentu menjadi konten baru yang diinginkan. Dalam PHP sendiri sudah disediakan library HTML DOM yang dapat diunduh terpisah dengan banyak keuntungan.

```

<?php
    require_once 'simple_html_dom.php';

    $html = file_get_html("http://hieppies.blogspot.com");
    $html = file_get_html("index.html");
    $html = file_get_html("/menu/index.html");

    $html = str_get_html("<html>
        <body>
            <p>Hi, Cantik!.</p>
            <p>Lagi ngapain?.</p>
        </body>
    </html>");
?>

```

**Kode Sumber 2.4-3 Syntax Penggunaan Library Simple HTML Dom Parser**

#### 2.4.4 Regular Expression (ReGex)

Merupakan teknik yang digunakan untuk pencocokan string teks, seperti karakter tertentu, kata-kata, atau pola karakter. ReGex memiliki 2 fungsi utama yaitu untuk mencari dan mengganti, mencari suatu pola tertentu dalam text lalu menggantinya menjadi pola yang lain. Dalam penggunaannya untuk grabbing situs website, ReGex biasanya dikombinasikan dengan fungsi `file_get_contents` dari PHP. Mengambil konten menggunakan fungsi `file_get_contents` sedangkan menganalisa hasil dan mencari string konten yang dibutuhkan menggunakan regular expression.

### 2.5 Pusat Data Scopus

Scopus merupakan database bibliografi yang meng-index abstrak dan sitasi/kutipan untuk artikel jurnal akademik. Scopus merupakan salah satu produk bibliograsi dari Elsevier. Scopus muncul sejak tahun 1995, kurang lebih 21 tahun Scopus meng-index jurnal dari para akademisi di seluruh dunia. Hingga saat ini kurang lebih terdapat 22.000 judul artikel jurnal dan lebih dari 5.000 penerbit [12].

Scopus meng-index artikel jurnal akademik dari berbagai bidang ilmu. Maka dari itu banyak publikasi pemeringkatan perguruan tinggi



lain seperti Scimago menggunakan jasa dari Scopus sebagai pusat database jurnal akademik. Scopus sendiri sangat berhati-hati dan selektif untuk meng-index artikel jurnal akademik. Tidak semua artikel jurnal di database lain seperti Google Scholar, ataupun Web of Science dapat ter-index oleh Scopus. Scopus hanya meng-index berdasarkan kualitas artikel jurnal tersebut.

Scopus juga menyediakan layanan untuk developer berupa Scopus API (Application Program Interfaces) sebagai sarana untuk mengembangkan informasi berdasarkan database Scopus, hanya saja untuk menggunakan layanan ini hanya untuk institusi yang berlangganan (*Subscribe*) saja [13]. Tidak semua orang dan institusi dapat menggunakan layanan ini. Keuntungan dari database Scopus ini adalah mempunyai atribut data dan informasi yang lengkap, diantaranya nama author, jumlah publikasi, jumlah sitasi, ORCID, dan grafik perkembangan keaktifan author [13]. Selain itu dukungan layanan API sangat lengkap dan juga kontennya dapat dipercaya.

## 2.6 Google Scholar

Google Scholar merupakan layanan database metadata publikasi yang dibuat oleh Google. Dalam Google Scholar memungkinkan pengguna melakukan pencarian materi berupa teks dalam berbagai format publikasi. Index Google Scholar mencakup jurnal online dan publikasi ilmiah sehingga semua jurnal online bisa tercover dalam Google Scholar [14]. Tujuan utama dari Google Scholar adalah menyusun paper seperti yang dilakukan peneliti, dengan memperhatikan kelengkapan teks, nama penulis yang menampilkan artikel, serta frekuensi penggunaan kutipan artikel dalam literatur akademis lainnya.

Dalam Google Scholar, seorang peneliti dapat mendaftarkan akun dirinya dalam sistem sebagai biodata untuk setiap paper yang dipublikasikan. Dalam setiap biodata penulis terdapat beberapa indikator yang cukup unik sebagai bahan acuan bahwa paper yang dipublikasikan oleh penulis tersebut berkualitas dan terpercaya atau tidak [15]. Hal ini dapat dilihat dari jumlah paper yang dipublikasikan penulis yang terindex oleh Google Scholar, jumlah sitasi (*citation*), nilai h-index, dan nilai i10-index [16]. Berikut penjelasan masing-masing indikator dalam Google Scholar:

### 1. Jumlah Publikasi

Merupakan jumlah paper yang dipublikasikan oleh penulis yang terindex oleh Google Scholar. Disini dapat diketahui juga tahun publikasi dan jumlah sitasi/kutipan terhadap paper tersebut.

Semakin banyak orang lain yang mengutip paper tersebut akan semakin tinggi peringkat posisi paper dalam daftar publikasi penulis tersebut. Jumlah publikasi ini dihitung sejak penulis aktif dalam Google Scholar.

2. Jumlah Sitasi

Merupakan jumlah total kutipan dari orang lain yang mengutip paper penulis tersebut. Semakin banyak orang yang mengutip paper penulis maka akan semakin banyak pula total sitasinya. Terdapat dua versi dalam jumlah sitasi pada Google Scholar, terdapat jumlah total (*all citation*) yang merupakan jumlah keseluruhan sejak penulis aktif dalam Google scholar, serta jumlah sejak 5 tahun terakhir (*Since 2012*) yang dapat dilihat bahwa kualitas sitasi dari paper tersebut baik atau tidak sejak 5 tahun terakhir.

3. Nilai H-Index

Merupakan ukuran index yang diusulkan oleh seorang fisikawan di Universitas California bernama Jorge E Hirsch pada tahun 1985. H-Index ini masih terus diteliti keakuratannya. H-Index sendiri adalah nilai index yang didapatkan dari jumlah paper sebanyak  $h$  paper penulis dengan jumlah kutipan untuk setiap *paper* minimal sama dengan nilai  $h$  tersebut [16].

4. Nilai i10-Index

Merupakan nilai index yang didapatkan dari jumlah 2 *paper* dari penulis yang dikutip minimal oleh 10 *paper* dari orang lain. Sehingga kondisi minimal dari kualitas *paper* tersebut dapat diketahui oleh cara ini.

## 2.7 *Science and Technology Index (SINTA)*

Science and Tecnology Index atau biasa disebut dengan SINTA merupakan pusat index, sitasi dan kepakaran terbesar di Indonesia. Sistem informasi riset berbasis website ini menawarkan akses cepat, mudah dan komprehensif untuk mengukur unjuk kerja peneliti, institusi dan jurnal di Indonesia. Sinta menyediakan benchmark dan analisis, identifikasi kekuatan riset tiap institusi, memperlihatkan kolaborasi penelitianm menganalisis tren penelitian dan direktori pakar. SINTA memberikan informasi mengenai tolok ukur institusi, kolaborasi, analisis penelitian serta direktori akademisi di Indonesia [17]. Beberapa informasi yang dapat digali dari database online SINTA [18], antara lain:

1. Jurnal yang diterbitkan oleh institusi pendidikan dan penelitian tinggi yang disajikan dalam bentuk peringkat.
2. Institusi pendidikan tinggi yang mempunyai kontribusi dalam penelitian, berisikan informasi jumlah author akademisi serta total sitasi yang dihasilkan dari masing-masing institusi pendidikan dan penelitian tinggi yang disajikan dalam bentuk peringkat.
3. Author akademisi yang mempunyai kontribusi dalam penelitian, berisikan informasi jumlah sitasi yang dimiliki, skor h-index, dan i10-index.
4. Detail informasi setiap institusi pendidikan tinggi:
  - a. Jumlah sitasi Google Scholar.
  - b. Jumlah dokumen Google Scholar.
  - c. Jumlah artikel pada Scopus database.
  - d. Jumlah non artikel pada Scopus database.
  - e. Jumlah sitasi dari Scopus.
  - f. Jumlah akademisi.
  - g. Daftar akademisi beserta kualitas sitasi (h-index, i10-index) dari Google Scholar dan Scopus database.
  - h. Jumlah jurnal artikel, book chapters, dan conference paper dari akademisi maupun institusi.
  - i. Skor penilaian menurut SINTA.

## **2.8 Pangkalan Data Pendidikan Tinggi (PDDIKTI)**

Pangkalan Data Pendidikan Tinggi atau lebih umum disebut PDDIKTI merupakan pusat kumpulan data penyelenggara pendidikan tinggi seluruh Indonesia. Kementerian Riset, Teknologi, dan Pendidikan Tinggi merupakan kementerian dalam pemerintahan Indonesia yang berfokus pada penyelenggaraan urusan bidang riset, teknologi dan pendidikan tinggi [19].

Data dalam PDDIKTI merupakan hasil sinkronisasi yang dikelola oleh masing-masing perguruan tinggi di Indonesia, sehingga akan lebih update dalam penyebaran informasi, serta terintegrasi secara nasional. Pangkalan data ini menjadi salah satu instrumen pelaksanaan penjaminan mutu. Data yang terangkum dalam PDDIKTI harus mudah diakses oleh stakeholder (lembaga akreditasi, pemerintah, masyarakat), sehingga menerapkan prinsip transparansi perguruan tinggi.

Beberapa data yang disediakan oleh PDDIKTI dalam sistem yang dapat dilihat [19], diantaranya:

1. Jumlah perguruan tinggi di Indonesia.
2. Jumlah program studi setiap perguruan tinggi.

3. Jumlah mahasiswa (jenis kelamin, kelompok bidang).
4. Jumlah dosen (jenis kelamin, ikatan kerja/status kepegawaian, pendidikan terakhir, jabatan akademik).

## 2.9 *Alexa Traffic Rank*

Alexa Traffic Rank merupakan sistem yang menyediakan fasilitas informasi tentang peringkat *webhost*, yang dilihat dari jumlah *traffic* pengunjung yang masuk kedalam situs *webhost* tersebut. Dengan kata lain Alexa Traffic Rank merupakan salah satu *tools* untuk *Search Engine Optimization* (SEO) website yaitu sitem untuk menganalisa dan mempengaruhi tingkat keterlihatan (*visibilitas*) sebuah situs pada mesin pencari [20].

Alexa Traffic Rank mempunyai nilai peringkat dari nilai terkecil yaitu nilai angka 1 (satu) hingga seterusnya (berjuta-juta) yang mana dalam perhitungan sistem di Alexa Traffic Rank apabila semakin kecil nilai Alexa Traffic Rank, nilai kepopuleran situs akan semakin bagus. Selain itu Alexa Traffic Rank juga menyediakan API untuk pengembangan (*developer*) perangkat lunak berbasis analisa SEO yang tidak berbayar (*free*).

## 2.10 *MOZ Analytics*

MOZ Analytics merupakan salah satu *tools* yang digunakan untuk mengetahui jumlah pencarian *webhost*. MOZ Analytics dikembangkan oleh SEOMoz yang berbasis di Seattle, Washington, Amerika Serikat. Salah satu product dari MOZ yang berfokus pada pemeringkatan dan analisa adalah MOZ Rank. MOZ Rank mengkuantifikasi kepopuleran link *webhost*. Halaman *webhost* mendapatkan skor MOZ Rank berdasarkan halaman lain yang terhubung dengan halaman *webhost* tersebut (ditautkan). Semakin banyak tautan dari halaman *webhost*, maka akan semakin tinggi skor dari MOZ Rank. Dengan kata lain MOZ Analytics dengan berbagai macam produk turunannya merupakan sebuah *tools* untuk *Search Engine Optimization* (SEO). MOZ Analytics mengembangkan API untuk *developer* SEO yang tidak berbayar (*free*).

## 2.11 **JAVA**

Java adalah bahasa pemrograman yang dapat dijalankan di berbagai komputer termasuk telepon genggam. Bahasa ini awalnya dibuat oleh James Gosling saat masih bergabung di Sun Microsystems saat ini merupakan bagian dari Oracle dan dirilis tahun 1995. Bahasa ini banyak mengadopsi sintaksis yang terdapat pada C dan C++ namun

dengan sintaksis model objek yang lebih sederhana serta dukungan rutin-rutin aras bawah yang minimal. Aplikasi-aplikasi berbasis java umumnya dikompilasi ke dalam p-code (bytecode) dan dapat dijalankan pada berbagai Mesin Virtual Java (JVM). Saat ini java merupakan bahasa pemrograman yang paling populer digunakan, dan secara luas dimanfaatkan dalam pengembangan berbagai jenis perangkat lunak aplikasi ataupun aplikasi [22]

## 2.12 PHP dan HTML

Hypertext Preprocessor atau yang biasa disebut dengan PHP merupakan bahasa script yang dapat disisipkan dalam kumpulan kode HTML [26]. PHP banyak dipakai dalam pemrograman website selain dengan HTML. PHP sendiri ditemukan oleh Rasmus Lerdorf pada tahun 1995, dan hingga sekarang PHP dikembangkan oleh vendor utama yaitu The PHP Group. Hingga saat ini PHP telah masuk dalam versi ke 7, atau biasa yang disebut dengan PHP 7. Kelebihan dari bahasa pemrograman PHP sebagai berikut:

1. Bahasa pemrograman yang tidak melakukan sebuah kompilasi dalam penggunaannya.
2. Web server yang mendukung PHP dapat ditemukan dimana-mana, diantaranya apache, IIS, Lighttpd, dan Xitami.
3. Dalam sisi pengembangan (*developer*) lebih mudah karena banyak organisasi maupun developer lain yang siap membantu mengembangkan variasi dari PHP sendiri.
4. Dalam sisi pemahaman, PHP adalah bahasa scripting yang paling mudah karena memiliki referensi yang baik dan banyak.
5. Open Source yang dapat digunakan di berbagai *Operating System* (OS) dan dapat dijalankan secara *runtime* melalui *console* serta dapat menjalankan perintah sistem.

## 2.13 Database Management System (DBMS) MySQL

MySQL merupakan sebuah perangkat lunak sistem manajemen basis (DBMS) data SQL. MySQL disebut database yang multithread, multi user dengan jumlah instalasi 6 juta kali di seluruh dunia yang membuat MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi GNU General Public License (GPL). MySQL sendiri dikembangkan oleh Oracle pada tahun 1995. MySQL sendiri bisa dijalankan disemua tipe sistem operasi [27].

MySQL adalah sebuah implementasi dari sistem manajemen basis data relasional RDBMS yaitu sistem basis data yang entitas utamanya terdiri dari tabel-tabel yang mempunyai relasi dari satu tabel

ke tabel yang lain. Setiap pengguna dapat secara bebas menggunakan MySQL, namun dengan batasan perangkat lunak tersebut tidak boleh dijadikan produk turunan yang bersifat komersial. MySQL sebenarnya merupakan turunan salah satu konsep utama dalam basis data yang telah ada sebelumnya, yaitu SQL (*Structured Query Language*). SQL adalah sebuah konsep pengoperasian basisdata, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis.

## 2.14 *Part-Of-Speech Tagger* (POS Tagger)

Merupakan teknik yang dapat menerapkan penanda tag dari dokumen teks (corpus) dalam beberapa bahasa untuk menjadi kelas kata berdasarkan tipe struktur kalimat dalam dokumen. Struktur kata ini seperti kata benda (*noun*), kata kerja (*verb*), kata sifat (*adjective*) dan sebagainya. POS Tagger disebut juga grammatical tagging. Proses menandai sebuah kata dalam sebuah teks (corpus) yang sesuai dengan bagian pembicaraan tertentu. Berdasarkan definisi konteks, yaitu hubungan dengan kata-kata yang berdekatan yang terkait dalam frase, kalimat atau paragraf [28].

POS Tagging pada zaman dahulu dilakukan menggunakan cara manual yaitu dengan menggunakan jasa ahli bahasa untuk memetakan kelas dari kata dalam corpus. Untuk saat ini POS Tagging dilakukan dalam konteks komputasi linguistik, menggunakan algoritma yang mengasosiasikan istilah diskrit, serta bagian pembicaraan yang tersembunyi (latent). Salah satu pengembangan POS Tagger yang populer saat ini adalah Stanford POS Tagger yang dibangun dalam bahasa Java. Struktur kelas kata POS Tagging digambarkan pada Tabel 2.15-1 [28].

**Tabel 2.14-1 Struktur Kelas kata POS Tagging**

Tag	Deskripsi	Contoh Penerapan
CC	Coordin, conjunction	And, but, or
CD	Cardinal number	One, two
DT	Determiner	A, the
EX	Existential 'there'	There
FW	Foreign word	Mea culpa
IN	Preposition/sub-conjunction	Of, in, by
JJ	Adjective	Yellow
JJR	Adj., comparative	Bigger

JJS	Adj., superlative	Wildest
LS	List item marker	1, 2, one
MD	Modal	Can, should
NN	Noun, sing. Or mass	Llama
NNS	Noun, plural	Llamas
NNP	Proper noun, sing.	IBM
NNPS	Proper noun, plural	Carolinas
PDT	Predeterminer	All, both
POS	Possessive ending	's
PRP	Personal pronoun	I, you, he
PRP\$	Possessive pronoun	Your, one's
RB	Adverb	Quickly, never
RBR	Adverb, comparative	Faster
RBS	Adverb, superlative	Fastest
RP	Participle	Up, off
SYM	Symbol	_, %, &
TO	"to"	to
UH	Interjection	Ah, oops
VB	Verb base form	Eat
VBD	Verb past tense	Ate
VBG	Verb gerund	Eating
VBN	Verb past participle	Eaten
VBP	Verb on-3sg pres	Eat
VBZ	Verb 3-sg pres	Eats
WDT	Wh-determiner	Which, that
WP	Wh-pronoun	What, who
WP\$	Possessive wh-	Whose
WRB	Wh-adverb	How, where
\$	Dollar sign	\$
#	Pound sign	#
"	Left quote	' or "
"	Right quote	' or "
(	Left parenthesis	[, (, {, <
)	Right parenthesis	], ), }, >
,	Comma	,
.	Sentence-final punc	. ! ?
:	Mid-sentence punc	: ; ... -- -

## 2.15 Bloom Taxonomy

Bloom Taxonomy merupakan sebuah cara untuk membagi cara belajar seseorang menjadi tiga ranah (*domain*) utama. Taksonomi ini dibuat untuk tujuan pendidikan. Dibuat pertama kali oleh Benjamin S. Bloom pada tahun 1956. Tiga ranah (*domain*) utama diantaranya [29]:

1. Cognitive Domain (ranah kognitif) berisi mengenai perilaku yang menekankan pada aspek intelektual, seperti pengetahuan, pengertian, dan keterampilan berpikir.
2. Affective Domain (ranah afektif) berisi mengenai perilaku yang menekankan pada aspek perasaan dan emosi, seperti minat, sikap, apresiasi, dan cara penyediaan diri.
3. Psychomotor Domain (ranah psychomotor) berisi mengenai aspek ketrampilan motorik seperti tulisan tangan, mengetik, dan mengoperasikan mesin.

Dari ketiga domain diatas, domain yang paling berpengaruh terhadap hasil intelektual adalah domain kognitif. Domain ini selanjutnya dibagi kedalam 6 level, diantaranya [29]:

1. Knowledge  
Menunjukkan materi yang telah dipelajari sebelumnya dengan mengingat fakta, istilah, konsep dasar dan jawaban.
2. Comprehension  
Menunjukkan pemahaman tentang fakta dan gagasan dengan mengatur, membandingkan, menerjemahkan, menafsirkan, memberikan deskripsi dan mengemukakan gagasan utama.
3. Application  
Memecahkan masalah dengan menerapkan pengetahuan, fakta, teknik dan peraturan yang diperoleh dengan cara yang berbeda.
4. Analysis  
Memeriksa dan memecahkan informasi menjadi beberapa bagian dengan mengidentifikasi motif atau sebab, serta membuat kesimpulan dan menemukan bukti untuk mendukung generalisasi.
5. Synthesis  
mengkompilasi informasi bersama-sama dengan cara yang berbeda dengan menggabungkan elemen dalam pola baru atau mengusulkan solusi alternatif.
6. Evaluation  
Menyajikan dan mempertahankan pendapat dengan membuat penilaian tentang informasi, validitas gagasan atau kualitas kerja berdasarkan seperangkat kriteria.



### 2.16 LSA (*Latent Semantic Analysis*)

Pendekatan PLSA, sebagian besar terpengaruh oleh metode yang biasa disebut Latent Semantic Analysis (LSA). Dalam LSA, kemunculan kata-kata dalam masing-masing dokumen dipetakan menggunakan Singular Value Decomposition (SVD) (Landauer, Foltz, & Laham, 1998). Proses dekomposisi ini memecah suatu matriks  $X$  berukuran  $m \times n$  menjadi 3 buah matriks  $U$ ,  $S$ , dan  $V^T$  yang masing-masing berukuran  $m \times m$ ,  $m \times n$ , dan  $n \times n$ .  $U$  adalah matriks dengan kolom ortonormal yang nilai-nilainya diperoleh dari 1 dibagi eigen value dan dikali eigen vektor dari  $X$ .  $S$  adalah matriks berukuran  $m \times n$  yang unsur-unsur diagonalnya merupakan nilai akar eigen value dari  $X$ .  $V^T$  adalah kolom ortonormal yang mengandung nilai-nilai eigen vektor dari  $X$ .

Contoh masukan dokumen untuk perhitungan LSA dapat dilihat pada Tabel 2.15.1 Dalam tabel tersebut terdapat 9 dokumen yang kemudian akan dilakukan perhitungan terhadap kata-kata dalam masing-masing dokumennya seperti yang terlihat pada Tabel 2.4.

**Tabel 2.15.1 Daftar Dokumen Masukan untuk Perhitungan LSA**

<b>Nama Dokumen</b>	<b>Isi Dokumen</b>
d1	Human machine interface for ABC computer application
d2	A survey of user opinion of computer system response time
d3	The EPS user interface management system
d4	System and human system engineering testing of EPS
d5	Relation of user perceived response time to error measurement
d6	The generation of random, binary, ordered trees
d7	The intersection graph of paths in trees
d8	Graph minors IV : Widhts of trees and well-quasi-ordering
d9	Graph minors: A survey

**Tabel 2.15.2 Daftar Frekuensi Kata pada Perhitungan LSA**

[illegible]

minors	0	0	0	0	0	0	0	1	1
--------	---	---	---	---	---	---	---	---	---

Pada Tabel 2.4. terdapat 9 dokumen pada barisnya dan 12 kata pada kolomnya dimana masing-masing angka di dalamnya menunjukkan nilai frekuensi kata tersebut dalam masing-masing dokumen.

Dari hasil perhitungan pada Tabel 2.4 kemudian dibentuk sebuah matriks X untuk masukan pada perhitungan LSA.

1	0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	1	1	0	1	0	0	0	0
0	1	1	2	0	0	0	0	0
0	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0	0
0	1	0	0	0	0	0	0	1
0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	1	1

**Gambar 2.15.1 Matriks X (Kemunculan Kata Dalam Masing-Masing Dokumen)**

Dari matriks pada Gambar 2.15.1 tersebut selanjutnya dapat dipetakan menggunakan SVD seperti pada Gambar 2.15.2, Gambar 2.15.3 dan Gambar 2.15.4

-0.2214	-0.1132	0.2890	-0.4148	-0.1063	-0.3410	-0.5227	0.0605	0.4067	-0.0759
-0.1976	-0.0721	0.1350	-0.5522	0.2818	0.4959	0.0704	0.0099	0.1089	-0.0339
-0.2405	0.0432	-0.1644	-0.5950	-0.1068	-0.2550	0.3022	-0.0623	-0.4924	0.1098
-0.4036	0.0571	-0.3378	0.0991	0.3317	0.3848	-0.0029	0.0004	-0.0123	-0.1517
-0.6445	-0.1673	0.3611	0.3335	-0.1590	-0.2065	0.1658	-0.0343	-0.2707	-0.1098
-0.2650	0.1072	-0.4260	0.0738	0.0803	-0.1697	-0.2829	0.0161	0.0539	-0.5750
-0.2650	0.1072	-0.4260	0.0738	0.0803	-0.1697	-0.2829	0.0161	0.0539	0.7267
-0.3008	-0.1413	0.3303	0.1881	0.1148	0.2722	-0.0330	0.0190	0.1653	0.2954
-0.2059	0.2736	-0.1776	-0.0324	-0.5372	0.0809	0.4669	0.0363	0.5794	0.0000

**Gambar 2.15.2 Matriks Hasil SVD ( $U$ )**

3.3409	0	0	0	0	0	0	0	0	0
0	2.5417	0	0	0	0	0	0	0	0
0	0	2.3539	0	0	0	0	0	0	0
0	0	0	1.6445	0	0	0	0	0	0
0	0	0	0	1.5048	0	0	0	0	0
0	0	0	0	0	1.3064	0	0	0	0
0	0	0	0	0	0	0.8459	0	0	0
0	0	0	0	0	0	0	0.5601	0	0
0	0	0	0	0	0	0	0	0.3637	0

**Gambar 2.15.3 Matriks Hasil SVD ( $S$ )**

-0.1974	-0.0559	0.1103	-0.9498	0.0457	-0.0766	-0.1773	0.0144	0.0637
-0.6060	0.1656	-0.4973	-0.0286	-0.2063	-0.2565	0.4330	-0.0493	-0.2428
-0.4629	-0.1273	0.2076	0.0416	0.3783	0.7244	0.2369	-0.0088	-0.0241
-0.5421	-0.2318	0.5699	0.2677	-0.2056	-0.3689	-0.2648	0.0195	0.0842
-0.2795	0.1068	-0.5054	0.1500	0.3272	0.0348	-0.6723	0.0583	0.2624
-0.0038	0.1928	0.0982	0.0151	0.3948	-0.3002	0.3408	-0.4545	0.6198
-0.0146	0.4379	0.1930	0.0155	0.3495	-0.2122	0.1522	0.7615	-0.0180
-0.0241	0.6151	0.2529	0.0102	0.1498	0.0001	-0.2491	-0.4496	-0.5199
-0.0820	0.5299	0.0793	-0.0246	-0.6020	0.3622	-0.0380	0.0696	0.4535

**Gambar 2.15.4 Matriks Hasil SVD ( $V^T$ )**

Dari hasil matriks SVD tersebut, dapat dipetakan menjadi 2 dimensi sehingga didapatkan gambar yang dapat menunjukkan kedekatan antar kata dan dokumen tersebut [31].

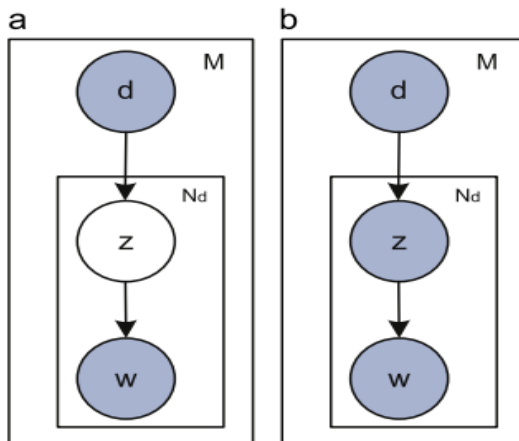
### **2.17 *pLSA ( Probabilistic Latent Semantic Analysis)***

Sebelum kemunculan Probabilistic Latent Semantic Analysis (PLSA) pada tahun 1999, 1 (satu) tahun sebelumnya, Landauer, Foltz dan Laham (1998) menciptakan satu teknik yang dinamakan Latent Semantic Analysis (LSA). LSA adalah sebuah teknik statistik terotomasi untuk membandingkan kesamaan semantik dari beberapa kata atau beberapa dokumen. LSA bukanlah sebuah program yang tradisional dari natural language processing ataupun artificial intelligence. Teknik ini digunakan dalam menganalisis dokumen untuk menemukan arti atau konsep dari dokumen tersebut. LSA lahir karena didasarkan pada pemikiran bahwa syntax dan style saja tidak mencukupi untuk menilai sebuah esai. Yang menjadi kesulitan mendasar adalah ketika kita hendak membandingkan kata-kata untuk menemukan dokumen yang relevan. Sebenarnya yang dibandingkan adalah arti atau konsep di balik kata-kata tersebut. Metode LSA melakukan mapping dari kata ataupun dokumen menjadi sebuah concept space dan perbandingan dilakukan pada space ini. Concept space tersebut atau yang lebih sering disebut sebagai latent semantic space merupakan hasil mapping dari matriks dimensi tinggi menjadi dimensi yang lebih kecil. Meskipun dalam dimensi yang lebih kecil, matriks tersebut merupakan matriks yang merepresentasikan isi dari keseluruhan dokumen. Ciri khas dari LSA adalah teknik yang dinamakan Singular Value Decomposition (SVD). SVD digunakan untuk melakukan dekomposisi matriks setelah diberikan pembobotan untuk kemudian diukur kesamaannya dengan data yang akan diujicobakan.

Permasalahan yang timbul dalam penggunaan metode LSA ini adalah adanya faktor polysemi dalam pengelompokkan kata (Hofmann, 2001). Permasalahan polysemi pada kata dapat diatasi dengan menggunakan varian dari LSA yang dikenal sebagai Probabilistic Latent Semantic Analysis (PLSA). Latent class Metode PLSA pada dasarnya merupakan model campuran dari model latent class dengan kata lain model latent class untuk data

teks, PLSA sendiri merupakan salah satu model based clustering yang bertujuan untuk membentuk cluster berdasarkan model peluang statistik, berbeda dengan metode cluster yang konvensional metode ini dapat dievaluasi berdasarkan ukuran statistik tertentu. Keutamaan metode PLSA sendiri dapat mereduksi dimensi matriks term yang terbentuk dari sekekumpulan teks yang direpresentasikan dalam sebuah variabel latent, sehingga ukuran dimensi matriks kemunculan term pada metode LSA dapat direduksi (Hofmann, 1999) [31].

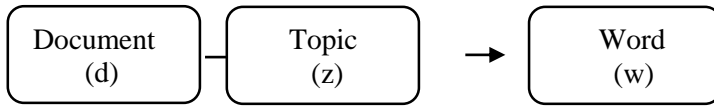
Supervised Probablistic Latent Semantic Analysis (sPLSA) digunakan untuk menghitung probabilitas pada sebuah kata-kata dan dokumen. Supervised PLSA dapat digunakan untuk mengidentifikasi kata-kata dengan beberapa arti dan pemetaan kata-kata dalam berbagai topik. Hubungan antara dokumen, topik, dan kata [35].



**Gambar 2.15 Grafik Perbedaan PLSA dan sPLSA**

Dari Grafik diatas merupakan grafik pLSA (a) dan SpLSA (b). Node diatas menunjukkan variabel secara acak. Node gelap

merupakan variabel yang dapat diketahui dan yang tidak gelap adalah variabel yang tidak dapat diketahui [34].



PLSA biasanya digunakan dalam aplikasi dari *Information Retrieval* atau *Natural Language Processing*. PLSA juga disebut model statistik (model aspek) untuk menemukan pola dalam dokumen teks sehingga lebih mudah untuk menghubungkan konteks dengan setiap kata yang muncul pada dokumen. Dengan proses pemodelan, topik atau konteks akan diperoleh dari teks asli dokumen tanpa deskripsi dokumen sebelumnya [34].

PLSA digunakan untuk mengklasifikasikan kata-kata menjadi topik yang belum diketahui (laten). Jadi, setiap dokumen yang dikelompokkan berdasarkan topik. algoritma ini adalah sebagai berikut: kita menentukan jumlah topik (z) kemudian menginisialisasi parameter probabilitas:  $P(z)$  adalah probabilitas topik,  $P(d|z)$  adalah dokumen probabilitas yang berisi topik,  $P(w|z)$  probabilitas kata yang terdapat dalam topik yang acak. Perhitungan kata pada sebuah dokumen dijelaskan dalam rumus dibawah ini:

For all  $k \leq n_j$ ,

$P(w_j, z_k) = \frac{n_{j,k}}{n_k}$  Sebagai inisialisasi  $P(w|z)$  dan inisialisasi secara acak dari  $P(z|d)$

$$P(d_i, w_j) = \sum_{k=1}^K P(Z_k)P(d_i|Z_k)P(W_j|Z_k) \quad (1)$$

Untuk menghitung manual bisa menggunakan rumus dibawah ini

$$p(wkj.ci) = \frac{f(wkj.ci) + 1}{f(ci) + |W|}$$

Keterangan:

$p(w_{kj}.ci)$  = probabilitas data

$f(w_{kj}.ci)$  = frekuensi kemunculan term pada dokumen

$f(ci)$  = frekuensi jumlah term pada dokumen

$|W|$  = jumlah term

Langkah selanjutnya adalah menghitung probabilitas untuk setiap parameter menggunakan *Expectation Maximization* dengan dua langkah, yaitu E langkah dan M langkah. E Langkah digunakan untuk menghitung probabilitas topik dalam dokumen dan dapat dilihat pada.

$$P(z_k | d_i, w_j) = \frac{P(w_j | z_k) P(z_k | d_i)}{\sum_{k=1}^K P(w_j | z_l) P(z_l | d_i)}$$

(2)

Langkah M digunakan untuk memperbaharui nilai parameter dan dapat dilihat di (3) dan (4).

$$P(w_{kj} | z_k) = \frac{\sum_{i=1}^N n(d_i | w_j) P(z_k | d_i w_j)}{\sum_{k=1}^K \sum_{m=1}^K n(d_i | w_m) P(z_k | d_i w_m)} \quad (3)$$

$$P(z_k | d_i) = \frac{\sum_{j=1}^N n(d_i | w_j) P(z_k | d_i w_j)}{n(d_i)} \quad (4)$$

Hasil perhitungan PLSA adalah probabilitas dari kata dalam topik dan kemungkinan topik dalam dokumen [35]. Tujuan dari menggunakan *Expectation Maximization* akan mengusahakan untuk memperoleh nilai error yang semakin kecil. Apabila nilai error masih tinggi, nilai bobot akan dioptimalkan supaya nilai error yang ada semakin kecil. Sehingga akan sangat memungkinkan bahwa apabila dilakukan semakin banyak training, maka topik serta probabilitas yang dibentuk akan semakin baik [36].



## 2.18 Fase Teks Mining (Metode pLSA) Studi Kasus Pemingkatan Perguruan Tinggi

Terdapat 6 tahap utama dalam alur metode *Probabilistic Latent Semantic Analysis (pLSA)*, dari dokumen mentah hingga menjadi bobot final reputasi akademik perguruan tinggi, diantaranya sebagai berikut:

### 1. Tahap preprocessing

Tahap preprocessing merupakan proses untuk membersihkan kata (*term*), dimana kumpulan dokumen abstrak paper yang terdiri dari banyak dokumen akan dikelola diambil kata bakunya saja. Kata baku menunjukkan komposisi keberadaan kata (*term*) dalam dokumen [22]. Tahap preprocessing terdiri dari beberapa sub-proses diantaranya:

#### 1.1. Case folding

Merupakan proses merubah huruf bertipe besar (*uppercase*) menjadi bertipe kecil (*lowercase*) dalam sebuah corpus dokumen.

#### 1.1. Tokenisasi

Merupakan proses pemecahan kumpulan kalimat menjadi kata (*term*) yang berdiri sendiri-sendiri (satu suku kata).

#### 1.2. Remove punctuation

Merupakan proses penghapusan karakter yang dapat disesuaikan pengguna. Pada umumnya karakter yang dihapus adalah karakter non-huruf (tanda baca - angka).

#### 1.3. Stopwords

Merupakan proses penghapusan kata (*term*) yang tidak begitu penting keberadaanya didalam dokumen. Sehingga dapat memaksimalkan keberadaan kata yang penting.

#### 1.4. Stemming

Merupakan proses pemotongan imbuhan (*prefix*, *suffix*, *infix*) dari sebuah kata (*term*) dalam sebuah dokumen.

Hasil dari tahap preprocessing berupa kumpulan kata baku dan telah bersih kemudian hasil kata ini akan menjadi bahan pairing dalam pencocokan Term Frequency (TF) untuk proses Menentukan asumsi label awal dokumen.

## 2. Tahap memperkaya corpus TKT

Corpus TKT merupakan dokumen kumpulan kata kunci (*keyword*) yang merepresentasikan kelas dari *term* pada dokumen dataset untuk masuk kedalam topik tingkat kesiapterapan teknologi (TKT) yang mempunyai 9 kelas. Corpus TKT mengadopsi corpus Taxonomy Bloom yang terdiri dari 6 kelas [29]. Untuk dapat digunakan sebagai corpus dalam teks mining metode PLSA, kelas corpus dalam Taxonomy Bloom perlu dipecah menjadi 9 kelas sesuai kelas TKT [5]. Breakdown menggunakan pengurutan aspek konten dari setiap *keyword*, dimana *keyword* yang mempunyai aspek tipe dasar berada pada level bawah dan aspek tipe penerapan berada pada level atas.

Selain menggunakan breakdown terhadap 6 corpus taxonomy Bloom menjadi 9 corpus TKT, dalam Tugas Akhir ini dilakukan juga pengembangan corpus dari beberapa cara. Cara pertama mengembangkan berdasarkan adopsi teknik POS Tagging dokumen abstrak. POS tagging akan memecah sebuah kalimat untuk diambil kelas tipe kata seperti verb, noun, adjective, adverb dan sebagainya [28]. Sebelum melakukan POS Tagging, dokumen abstrak paper akan dilakukan pencocokan terhadap *keyword* di masing-masing corpus TKT. setelah melakukan pencocokan dan ditemukan kecocokan, selanjutnya akan dicek terlebih dahulu kata (*term*) tersebut tergolong pada kelas kata apa dari prinsip POS Tagging, setelah diketahui kelas tipe kata kemudian akan mengakuisisi kata (*term*) lain yang berdekatan dengan kata (*term*) tersebut dengan jangkauan akuisisi beragam, mulai dari akuisisi 1 hingga 10. Dalam percobaan selanjutnya acuan menggunakan akuisisi 1. Kelas kata (*term*) hasil akuisisi akan

masuk pada corpus yang sama dengan kata yang dicocokkan sebelumnya.

Untuk cara kedua mengembangkan berdasarkan melihat sinonim dari setiap kata yang ada pada 9 corpus TKT. Setiap kata (term) akan dicari sinonimnya menggunakan library WordNet dalam bahasa pemrograman python [24]. Terdapat 4 tipe percobaan dalam menghasilkan corpus baru dengan mengacu pada pengembangan corpus dari segi sinonim, percobaan pertama dengan melihat sinonim level 1 dari setiap kata yang ada pada 9 corpus TKT. sinonim level 1 didapatkan dari satu nilai similaritas paling tinggi antara sinonim. Percobaan kedua dengan melihat sinonim level 2 dari setiap kata. Sinonim level 2 didapatkan dari dua nilai similaritas tertinggi. Percobaan 3 menggunakan prinsip yang sama hanya saja merubahnya menjadi tiga nilai similaritas tertinggi. Sedangkan yang terakhir adalah menggunakan sinonim level all, dimana dalam percobaan ini akan diambil semua sinonim dari setiap kata yang ada pada 9 corpus TKT. hasil dari pencarian sinonim akan ditambahkan pada corpus TKT final.

Dalam Tugas Akhir ini disimpulkan menggunakan proses memperkaya corpus TKT dengan cara mengembangkan dari segi sinonim pada percobaan sinonim level 3 dan all sebagai perbandingan. Karena berdasar hasil analisa akhir menunjukkan bahwa distance atau bisa disebut hasil akhir dari kedua percobaan ini lebih bagus dari percobaan lainnya.

### **3. Tahap penentuan asumsi topik awal menggunakan Term Frequency (TF)**

Tahap ini dilakukan dengan menghitung kemunculan term paling banyak dari dokumen abstrak *paper* pada 9 corpus TKT final [37]. 9 corpus TKT final yang digunakan berasal dari beberapa percobaan dalam tahap sebelumnya, diantaranya menggunakan corpus TKT hasil memperkaya menggunakan teknik POS Tagging akuisisi 1, percobaan sinonim WordNet

1, percobaan sinonim WordNet 2, percobaan sinonim WordNet 3, dan percobaan sinonim WordNet all. Dari semua percobaan ini disimpulkan bahwa menggunakan percobaan 3 dan all lebih bagus dalam menghasilkan distance hasil akhir.

Dokumen yang mempunyai kecocokan berdasarkan term frequency (TF) akan dicatat jumlah term yang sesuai, dan diurutkan berdasarkan jumlah kata yang paling tinggi hingga yang paling rendah. Kecocokan ini menandakan bahwa dokumen relevan berdasarkan keyword dengan ke-9 corpus TKT. 3 kecocokan teratas antara dokumen dan corpus TKT akan digunakan sebagai asumsi label awal dokumen untuk masukan tahap inferensi Gibbs Sampling PLSA.

#### **4. Tahap generative model PLSA**

Tahap ini merupakan proses ini untuk menghasilkan probabilitas akhir setiap label yang menjadi topik final. Masukan pada tahap ini berupa asumsi label topik yang diharuskan lebih dari satu label beserta dokumen abstrak paper. Asumsi label topik ini akan diproses menggunakan threshold, iterasi dan ketentuan variabel masukan lainnya yang ditentukan oleh user. Penggunaan metode PLSA berguna untuk memastikan asumsi label mempunyai probabilitas tertinggi agar berada pada satu label saja dari sebuah dokumen uji. Hasil keluaran proses PLSA berupa probabilitas term dari dokumen abstrak dengan asumsi label topik, serta probabilitas dokumen abstrak dengan asumsi label topik. Dalam Tugas Akhir ini, keluaran yang berperan penting untuk proses selanjutnya adalah probabilitas dokumen abstrak dengan label topik.

#### **5. Tahap penentuan label topik final dan pembobotan**

Pada tahap ini telah didapatkan topik final dari asumsi label topik yang dijadikan batasan uji Gibbs Sampling. Masing-masing asumsi label topik ini berguna untuk proses pembobotan. Setiap asumsi label topik hasil keluaran PLSA beserta probabilitas akan dihitung sebagai masukan tahap

penghitungan skor final. Hasil penghitungan ini akan mencerminkan skor kriteria reputasi akademik dari perguruan tinggi.

## **2.19 Akreditasi Perguruan Tinggi**

Akreditasi perguruan tinggi merupakan salah satu cara untuk menilai kualitas perguruan tinggi secara keseluruhan, yang bertujuan untuk mengetahui komitmen perguruan tinggi terhadap kapasitas dan efektifitas pendidikan yang diselenggarakan, yang didasarkan pada standar akreditasi yang telah ditetapkan oleh badan akreditasi nasional [38].

Di Indonesia pihak yang berwenang untuk melakukan akreditasi institusi perguruan tinggi adalah Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Akreditasi ini dilakukan oleh tim asesor yang turun langsung kepada perguruan tinggi terkait untuk dinilai dari beberapa hal. Berikut adalah poin-poin penting dalam penilaian akreditasi institusi oleh BAN-PT [38].

1. Menilai kesesuaian visi, misi, tujuan dan sasaran, serta strategi pencapaian dalam perguruan tinggi dengan kondisi yang sudah ada sekarang.
2. Menilai tata pamong (kepengurusan), kepemimpinan, sistem pengelolaan, dan penjaminan mutu.
3. Menilai kualitas mahasiswa dan lulusan.
4. Menilai sumber daya manusia (tenaga pengajar dan administrasi).
5. Menilai kesesuaian kurikulum, model pembelajaran, serta suasana akademik.
6. Menilai kebutuhan pembiayaan, sarana dan prasarana, serta sistem informasi dalam perguruan tinggi.
7. Menilai tentang penelitian, pelayanan/pengabdian masyarakat, serta kerjasama.

Penilaian akreditasi ini dilakukan dengan cara survey langsung terhadap perguruan tinggi, sehingga data yang didapatkan dapat objektif terhadap perguruan tinggi tersebut. Akreditasi yang

dilakukan oleh BAN-PT tersebut mempunyai 2 jenis, yaitu akreditasi institusi dan akreditasi profesi.

Akreditasi institusi yang merupakan penilaian kualitas institusi dengan memperhatikan poin-poin penilaian diatas. Sedangkan akreditasi profesi merupakan penilaian kualitas terhadap bidang studi/departemen/jurusan dalam perguruan tinggi tersebut, yang mana akan menentukan kualitas untuk akreditasi institusi juga. Berikut adalah score penilaian akreditasi BAN-PT.

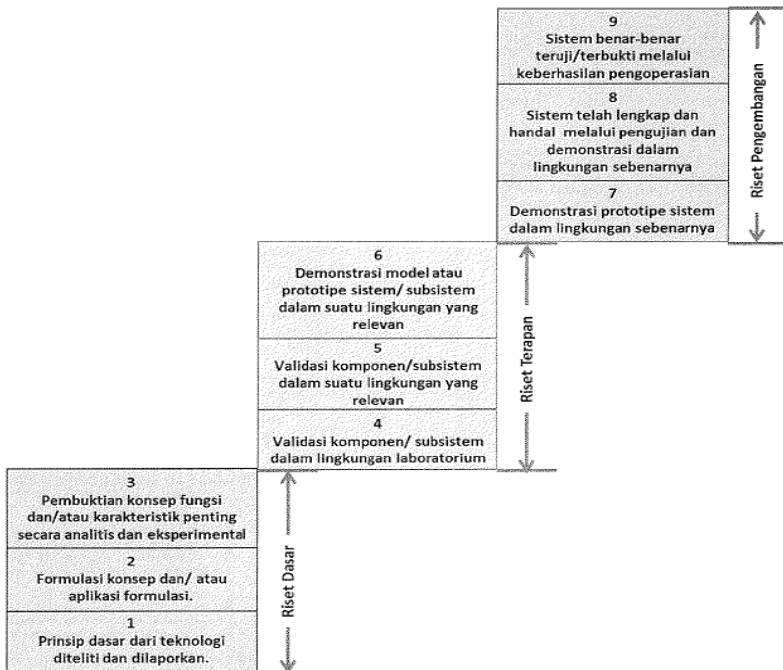
- a) Nilai 601 – 700, mendapatkan peringkat A
- b) Nilai 501 – 600, mendapatkan peringkat B
- c) Nilai 401 – 500, mendapatkan peringkat C
- d) Nilai 0 – 400, mendapatkan peringkat N/A (not applicable)

Dalam Tugas Akhir ini, salah satu kriteria penilaian yang menjadi acuan adalah kriteria reputasi employer. Dengan menggunakan kriteria akreditasi perguruan tinggi oleh BAN-PT yang dijelaskan pada poin penilaian kualitas mahasiswa dan lulusan sudah mengandung kriteria reputasi employer yang dibutuhkan.

## **2.20 *Tingkat Kesiapterapan Teknologi (TKT)***

Tingkat Kesiapterapan Teknologi atau yang biasa disebut dengan TKT merupakan tingkat kondisi kematangan atau kesiapterapan suatu hasil penelitian (*research*) dan pengembangan prototype tertentu yang selanjutnya akan masuk pada tahap komersialisasi, yang diukur secara sistematis dengan tujuan untuk dapat dimanfaatkan oleh pengguna, baik oleh pemerintah, maupun masyarakat luas [5].

Setiap jenis penelitian (*research*) dan pengembangan prototype dikelompokkan menjadi 9 tingkatan dengan masing-masing memiliki indikator yang berbeda-beda. Seperti pada Gambar 2.22-1.



**Gambar 2.20-1 Hirarki Tingkat Kesiapterapan Teknologi (TKT)**

Berdasarkan hirarki tingkatan diatas, dapat dikelompokkan menjadi 3 kelompok tingkatan besar [5], diantaranya.

### 1. Riset dasar untuk TKT tingkat 1-3

Dalam kelompok ini, penelitian yang dimaksud adalah penelitian yang berfokus pada pembahasan metode, pengembangan teori, pembahasan gagasan produk. Penelitian ini akan bertujuan untuk membuat produk yang bermanfaat bagi masyarakat.

### 2. Riset terapan untuk TKT tingkat 4-6

Dalam kelompok ini, penelitian yang dimaksud adalah penelitian yang berfokus pada yang menghasilkan produk,

demonstrasi model atau prototype subsistem dalam lingkungan yang relevan, dan berfokus pada validasi dalam lingkungan laboratorium. Penelitian ini sudah membuat produk/prototype.

**3. Riset pengembangan untuk TKT tingkat 7-9**

Dalam kelompok ini, penelitian yang dimaksud adalah penelitian yang berfokus pada pengembangan produk yang sudah ada sehingga membuat riset pendukung baru untuk produk yang sudah ada.



*[Halaman ini sengaja dikosongkan]*

## **BAB III**

### **ANALISIS DAN PERANCANGAN**

Pada bab ini diuraikan tahap analisis kebutuhan dan perancangan perangkat lunak dalam Tugas Akhir. Tahap analisis membahas mengenai analisis kebutuhan yang menjadi dasar dari tahap perancangan sistem. Perancangan sistem direpresentasikan dengan diagram UML (*Unified Modelling Language*).

#### **3.1. Tahap Analisis**

Tahap analisis mendefinisikan kebutuhan yang akan dipenuhi dalam pembangunan perangkat lunak pemeringkatan perguruan tinggi. Terdapat beberapa bagian antara lain, deskripsi umum, aktor dalam sistem, spesifikasi kebutuhan sistem, kasus penggunaan, dan dataset pengujian sistem.

##### **3.1.1 Deskripsi Umum**

Pada Tugas Akhir ini dibangun perangkat lunak pemeringkatan perguruan tinggi di Indonesia untuk mengukur tingkat kesiapterapan teknologi (TKT) menggunakan metode *Probabilistic Latent Semantic Analysis (pLSA)*. Perangkat lunak yang dibangun berbasis pada aplikasi website.

Data masukan yang digunakan adalah data yang mewakili berbagai macam kriteria penilaian perguruan tinggi yang mengacu pada *QS World University Rankings*, diantaranya adalah dokumen abstrak perguruan tinggi yang akan diolah menggunakan proses teks mining, data hasil akreditasi institusi yang didapatkan dari data online BAN-PT, data rekap jumlah mahasiswa perguruan tinggi, data rekap jumlah tenaga pengajar dari berbagai spesifikasi (status ketenagakerjaan, pendidikan terakhir, dan jabatan akademik), data jumlah tenaga kerja Phd (S3) yang didapatkan dari database PDDIKTI, data jumlah sitasi akademisi perguruan tinggi, data kualitas sitasi yang dilihat dari skor h-index dan i10-index akademisi perguruan tinggi, data jumlah publikasi yang terindex oleh Scopus yang didapatkan dari database SINTA, jumlah backlink *webhost*, serta skor Page Authority dan Domain Authority

*webhost* perguruan tinggi yang didapatkan dengan analisa *webhost*. Data keluaran dari perangkat lunak ini merupakan informasi peringkat perguruan tinggi di Indonesia berdasarkan pembobotan pada data inputan yang digunakan sebagai penilaian posisi peringkat, selain itu juga menghasilkan detail informasi tambahan mengenai kualitas institusi, akademisi dan kemahasiswaan.

### 3.1.2 Aktor Dalam Sistem

Terdapat 1 aktor dalam sistem pemeringkatan perguruan tinggi ini, yaitu pengguna sistem perangkat lunak website (*End User*). Pengguna *End User* mempunyai hak akses dalam menentukan bobot masukan untuk pembobotan perguruan tinggi, melihat pemeringkatan berdasarkan metode dan tipe pemeringkatan lain, mencari perguruan tinggi berdasarkan kebutuhan, serta melihat detail informasi akademik perguruan tinggi. Pengguna *End User* sebagai pengguna tunggal dalam sirkulasi sistem perangkat lunak.

### 3.1.3 Spesifikasi Kebutuhan Sistem

Pada perangkat lunak pemeringkatan perguruan tinggi di Indonesia untuk mengukur tingkat kesiapterapan teknologi (TKT) menggunakan metode *Probabilistic Latent Semantic Analysis (pLSA)* ini mempunyai spesifikasi kebutuhan yang mencakup kebutuhan fungsional sistem. Kebutuhan fungsional berisikan proses-proses yang dibutuhkan dalam sistem dan harus dijalankan agar menghasilkan pemeringkatan yang akurat terhadap *Ground Truth* yang telah ditetapkan. Kebutuhan fungsional tersebut dideskripsikan dalam Tabel 1-1.

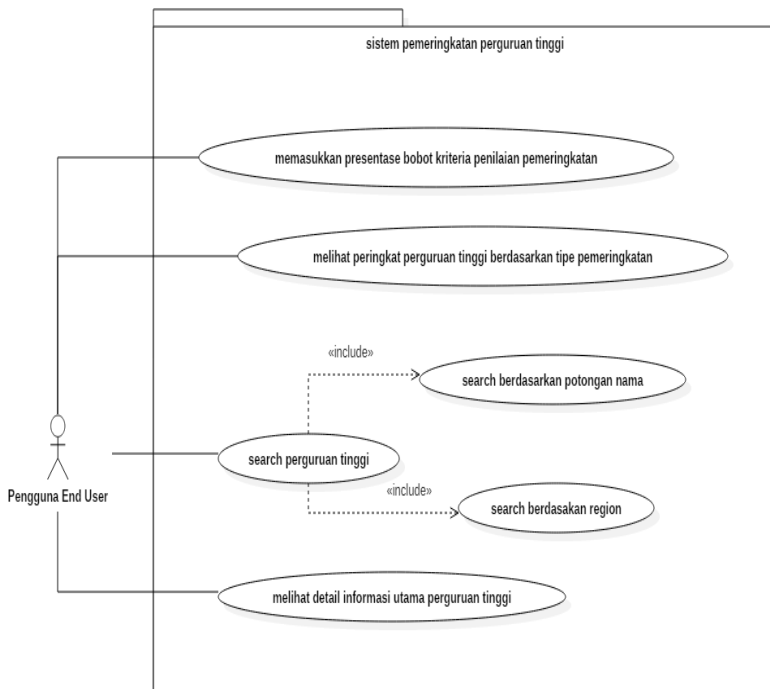
**Tabel 1-1 Daftar Kebutuhan Fungsional Sistem**

Kode Kebutuhan	Kebutuhan Fungsional	Deskripsi	Aktor
F-001	Memasukkan presentase bobot kriteria	Pengguna dapat memasukkan presentase bobot pemeringkatan sesuai dengan kebutuhan.	End User

	penilaian pemeringkatan		
<b>F-002</b>	Melihat peringkat perguruan tinggi berdasarkan tipe pemeringkatan	Pengguna dapat melihat peringkat perguruan tinggi dengan menggunakan bobot yang dipilih sesuai dengan tipe pemeringkatan.	End User
<b>F-003</b>	Mencari perguruan tinggi dari database berdasarkan potongan nama	Pengguna dapat mencari perguruan tinggi dengan memasukkan keyword berupa potongan nama perguruan tinggi dalam kotak pencarian search box.	End User
<b>F-004</b>	Mencari perguruan tinggi dari database berdasarkan region wilayah provinsi	Pengguna dapat mencari perguruan tinggi dengan memilih region wilayah provinsi yang disediakan dalam map peta Indonesia dalam 33 provinsi.	End User
<b>F-005</b>	Melihat detail informasi akademik perguruan tinggi	Pengguna dapat melihat informasi detail dari perguruan tinggi, yang dapat dikelompokkan menjadi 6 kategori, diantaranya kepopuleran webhost, mahasiswa, akreditasi, produktifitas institusi, kecukupan dosen pengajar, dan kualitas akademisi.	End User

#### 3.1.4 Kasus Penggunaan

Kasus penggunaan yang dibutuhkan pada sistem sesuai dengan analisa yang sudah dilakukan. Diagram kasus penggunaan yang mewakili kebutuhan sistem dapat dilihat pada Gambar 1-1 dan kode kasus penggunaan pada Tabel 1-2.



**Gambar 1-1 Diagram Kasus Penggunaan**

**Tabel 1-2 Daftar Kebutuhan Fungsional Sistem**

Kode Kasus Penggunaan	Kasus Penggunaan
UC-001	Memasukkan presentase bobot kriteria penilaian pemeringkatan
UC-002	Melihat peringkat perguruan tinggi berdasarkan tipe pemeringkatan
UC-003	Mencari perguruan tinggi berdasarkan potongan nama
UC-004	Mencari perguruan tinggi berdasarkan region wilayah provinsi

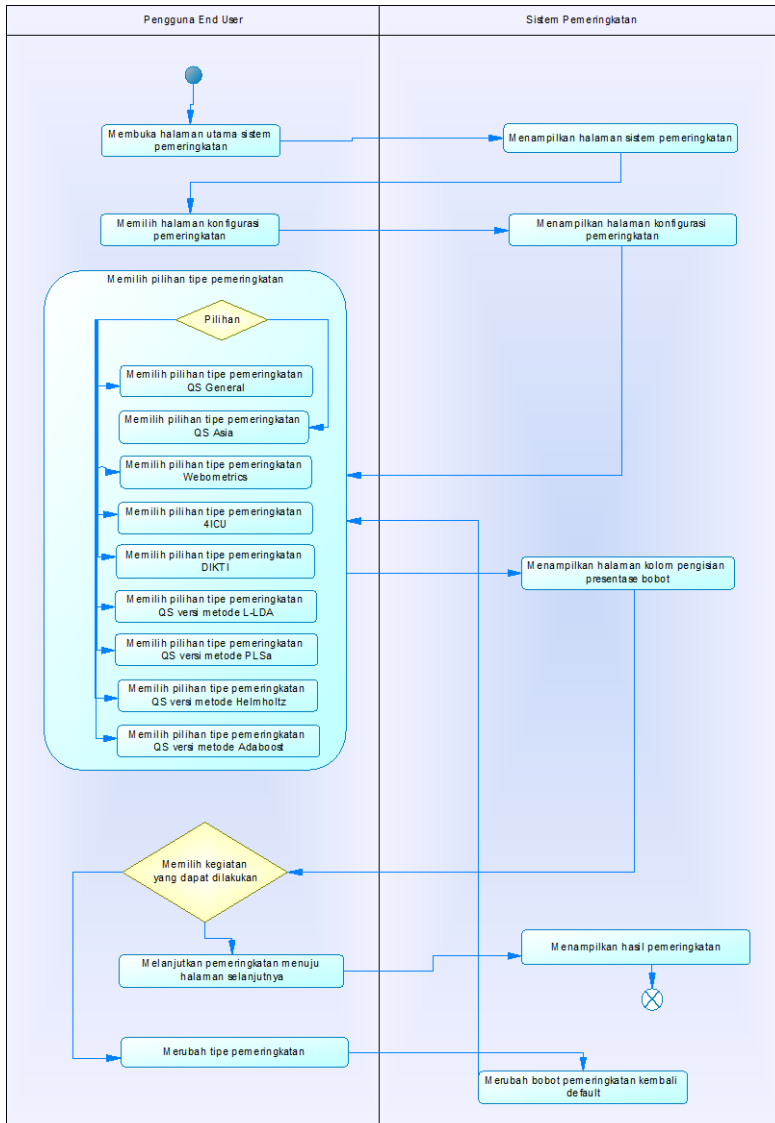
UC-005	Melihat detail informasi utama perguruan tinggi
--------	---

### 3.1.4.1 UC-001 Memasukkan Presentase Bobot Kriteria Penilaian Pemeringkatan

**Tabel 1-3 Case Memeasukkan Presentase Bobot Kriteria Penilaian Pemeringkatan**

<b>Nama</b>	Memasukkan Presentase Bobot Kriteria Penilaian Pemeringkatan
<b>Nomor</b>	UC-001
<b>Deskripsi</b>	Kasus penggunaan ini digunakan untuk menentukan bobot pemeringkatan pada tipe pemeringkatan versi metode PLSA dengan menentukan bobot pemeringkatan disetiap kriteria yang diinginkan oleh pengguna
<b>Tipe</b>	Fungsional
<b>Aktor</b>	Pengguna End User
<b>Kondisi Awal</b>	Pengguna sudah masuk dalam sistem
<b>Kondisi Akhir</b>	Sistem menampilkan halaman pemilihan kriteria dan tipe pemeringkatan yang sudah mempunyai bobot
<b>Alur Normal</b>	<ol style="list-style-type: none"> <li>1. Pengguna End User membuka halaman utama sistem pemeringkatan</li> <li>2. Sistem menampilkan halaman utama sistem pemeringkatan</li> <li>3. Pengguna memilih halaman konfigurasi pemeringkatan dengan melakukan scroll kursor kebawah</li> <li>4. Sistem menampilkan halaman konfigurasi pemeringkatan</li> <li>5. Pengguna End User memilih salah satu pilihan tipe pemeringkatan yang terdapat pada halaman konfigurasi pemeringkatan <ol style="list-style-type: none"> <li>a. Tipe pemeringkatan QS General</li> <li>b. Tipe pemeringkatan QS versi Asia</li> <li>c. Tipe pemeringkatan Webometrics</li> <li>d. Tipe pemeringkatan 4ICU</li> <li>e. Tipe pemeringkatan DIKTI</li> <li>f. Tipe pemeringkatan QS versi metode LLDA</li> <li>g. Tipe pemeringkatan QS versi metode PLSA</li> <li>h. Tipe pemeringkatan QS versi metode Helmholtz</li> <li>i. Tipe pemeringkatan QS versi metode Adaboost LDA</li> </ol> </li> </ol>

	6. Sistem menampilkan halaman kolom pengisian presentase bobot pada tipe pemeringkatan yang dipilih 7. Pengguna End User memilih kegiatan yang dapat dilakukan <ul style="list-style-type: none"> <li>j. Melanjutkan pemeringkatan menuju halaman selanjutnya</li> <li>k. Merubah tipe pemeringkatan</li> </ul>
<b>Alur Alternatif</b>	-



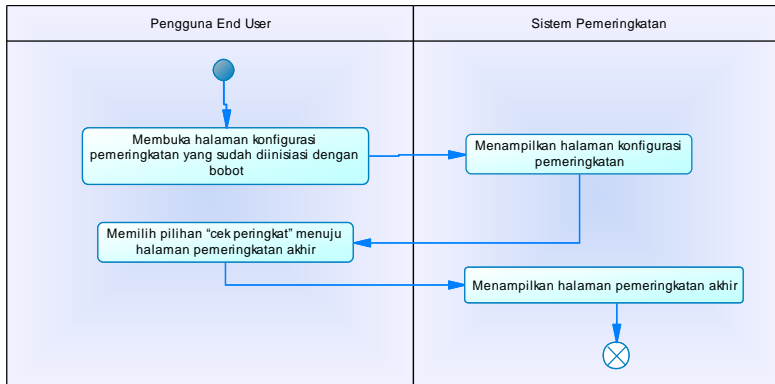
**Gambar 1-2 Diagram Aktivitas Use Case UC-001**



### 3.1.4.2 UC-002 Melihat Peringkat Perguruan Tinggi Berdasarkan Tipe Pemeringkatan

**Tabel 1-4 Use Case Melihat Peringkat Perguruan Tinggi Berdasarkan Tipe Pemeringkatan**

<b>Nama</b>	Melihat Peringkat Perguruan Tinggi Berdasarkan Tipe Pemeringkatan
<b>Nomor</b>	UC-002
<b>Deskripsi</b>	Kasus penggunaan ini digunakan untuk melihat pemeringkatan akhir perguruan tinggi
<b>Tipe</b>	Fungsional
<b>Aktor</b>	Pengguna End User
<b>Kondisi Awal</b>	Pengguna sudah masuk dalam sistem dan memasukkan bobot kriteria penilaian
<b>Kondisi Akhir</b>	Sistem menampilkan halaman akhir yang menampilkan peringkat perguruan tinggi sesuai dengan bobot dan pilihan tipe pemeringkatan yang diinisiasi sebelumnya
<b>Alur Normal</b>	<ol style="list-style-type: none"> <li>1. Pengguna End User membuka halaman konfigurasi pemeringkatan yang sudah diinisiasi dengan bobot</li> <li>2. Sistem menampilkan halaman konfigurasi pemeringkatan yang sudah diinisiasi dengan bobot</li> <li>3. Pengguna End User memilih pilihan “cek peringkat” menuju halaman pemeringkatan akhir</li> <li>4. Sistem menampilkan halaman pemeringkatan akhir</li> </ol>
<b>Alur Alternatif</b>	-



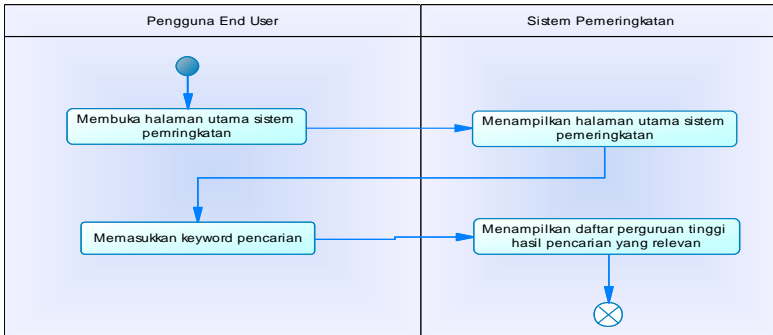
**Gambar 1-3 Diagram Aktifitas Use Case UC-002**

### 3.1.4.3 UC-003 Mencari Perguruan Tinggi Berdasarkan Potongan Nama

**Tabel 1-5 Use Case Mencari Perguruan Tinggi Berdasarkan Potongan Nama**

<b>Nama</b>	Mencari Perguruan Tinggi Berdasarkan Potongan Nama
<b>Nomor</b>	UC-003
<b>Deskripsi</b>	Kasus penggunaan ini digunakan untuk mencari informasi perguruan tinggi berdasarkan masukan berupa potongan nama perguruan tinggi pada kolom search pada halaman utama
<b>Tipe</b>	Fungsional
<b>Aktor</b>	Pengguna End User
<b>Kondisi Awal</b>	Pengguna sudah masuk dalam sistem
<b>Kondisi Akhir</b>	Sistem menampilkan halaman hasil pencarian berdasarkan keyword potongan nama yang dimasukkan dalam kolom search yang berisikan list perguruan tinggi yang relevan
<b>Alur Normal</b>	<ol style="list-style-type: none"> <li>1. Pengguna End User membuka halaman utama sistem pemeringkatan</li> <li>2. Sistem menampilkan halaman utama sistem pemeringkatan</li> <li>3. Pengguna End User memasukkan keyword pencarian berupa potongan nama pada kolom pencarian</li> </ol>

	4. Sistem menampilkan daftar perguruan tinggi hasil pencarian yang relevan dalam halaman result
<b>Alur Alternatif</b>	-



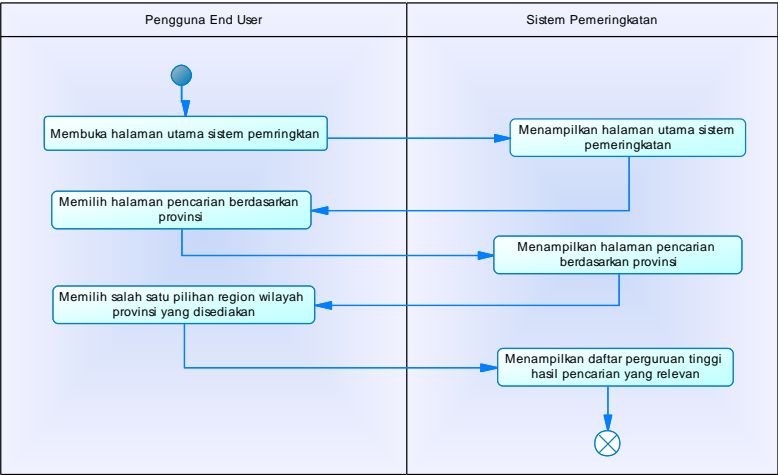
**Gambar 1-4 Diagram Aktifitas Use Case UC-003**

#### 3.1.4.4 UC-004 Mencari Perguruan Tinggi Berdasarkan Region Wilayah Provinsi

**Tabel 1-6 Use Case Mencari Perguruan Tinggi Berdasarkan Region Wilayah Provinsi**

<b>Nama</b>	Mencari Perguruan Tinggi Berdasarkan Region Wilayah Provinsi
<b>Nomor</b>	UC-004
<b>Deskripsi</b>	Kasus penggunaan ini digunakan untuk mencari informasi perguruan tinggi berdasarkan kategori region wilayah provinsi perguruan tinggi pada halaman pencarian provinsi
<b>Tipe</b>	Fungsional
<b>Aktor</b>	Pengguna End User
<b>Kondisi Awal</b>	Pengguna sudah masuk dalam sistem
<b>Kondisi Akhir</b>	Sistem menampilkan halaman hasil pencarian berdasarkan region wilayah provinsi yang dipilih pada map provinsi
<b>Alur Normal</b>	1. Pengguna End User membuka halaman utama sistem pemeringkatan

	<div>2. Sistem menampilkan halaman utama sistem pemeringkatan</div> <div>3. Pengguna End User memilih halaman pencarian berdasarkan provinsi dengan melakukan scroll kursor kebawah</div> <div>4. Sistem menampilkan halaman pencarian berdasarkan provinsi</div> <div>5. Pengguna End User memilih salah satu pilihan region wilayah provinsi yang disediakan</div> <div>6. Sistem menampilkan daftar perguruan tinggi yang berada pada region wilayah provinsi yang dipilih</div>
Alur Alternatif	-

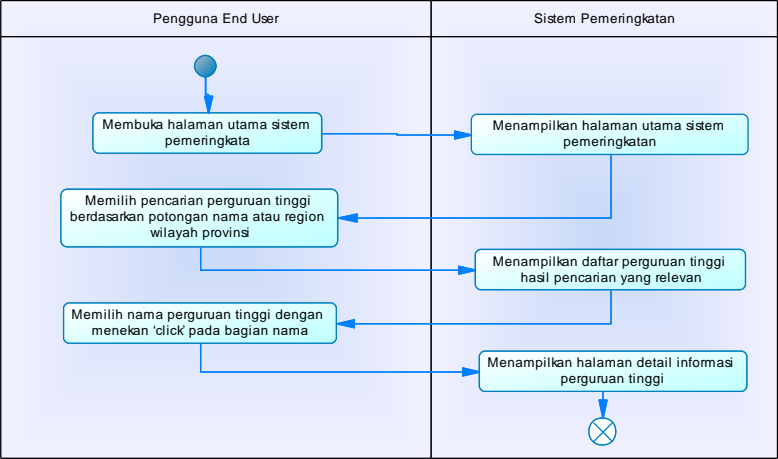


Gambar 1-5 Diagram Aktivitas Use Case UC-004

3.1.4.5 UC-005 Melihat Detail Informasi Utama Perguruan Tinggi

Tabel 1-7 Use Case Melihat Detail Informasi Utama Perguruan Tinggi	
Nama	Melihat Detail Informasi Utama Perguruan Tinggi
Nomor	UC-005

<b>Deskripsi</b>	Kasus penggunaan ini digunakan untuk melihat detail informasi yang dimiliki perguruan tinggi, meliputi kepopuleran webhost, mahasiswa, status akreditasi, produktifitas institusi, kecukupan dosen pengajar, dan kualitas akademisi serta dilengkapi dengan grafik tipe pemeringkatan yang dimiliki
<b>Tipe</b>	Fungsional
<b>Aktor</b>	Pengguna End User
<b>Kondisi Awal</b>	Pengguna sudah melakukan pencarian perguruan tinggi berdasarkan potongan nama atau region wilayah provinsi dan sistem menampilkan daftar perguruan tinggi
<b>Kondisi Akhir</b>	Sistem menampilkan halaman detail informasi yang mengandung banyak informasi, diantaranya kepopuleran webhost, mahasiswa, status akreditasi, produktifitas institusi, kecukupan dosen pengajar, dan kualitas akademisi serta dilengkapi dengan grafik tipe pemeringkatan yang dimiliki
<b>Alur Normal</b>	<ol style="list-style-type: none"> <li>1. Pengguna End User membuka halaman utama sistem pemeringkatan</li> <li>1. Sistem menampilkan halaman utama sistem pemeringkatan</li> <li>2. Pengguna End User memilih pencarian perguruan tinggi berdasarkan potongan nama atau region wilayah provinsi</li> <li>3. Sistem menampilkan daftar perguruan tinggi hasil pencarian yang relevan pada halaman pencarian</li> <li>4. Pengguna End User memilih nama perguruan tinggi dengan menekan 'click' pada bagian nama</li> <li>5. Sistem menampilkan halaman detail informasi perguruan tinggi</li> </ol>
<b>Alur Alternatif</b>	-



Gambar 1-6 Diagram Aktivitas Use Case UC-005







### 3.1.7 Label Kelas dalam Teks Mining Metode PLSA

Penggunaan metode PLSA yang bersifat supervised mengharuskan adanya keberadaan kelas atau bisa disebut juga dengan label. Label ini digunakan sebagai penanda pengetahuan dokumen masuk kedalam kondisi seperti apa. Kondisi dalam studi kasus ini adalah hirarki dari Tingkat Kesiapterapan Teknologi yang berjumlah 9 level. Berikut adalah daftar kelas yang digunakan dalam Tugas Akhir ini.

**Tabel 1-8 Penjelasan Setiap Kelas Label**

No.	Hirarki TKT	Kelas/Label	Keterangan
1	Level 1	T1	Prinsip dasar dari teknologi diteliti dan dilaporkan.
2	Level 2	T2	Formulasi konsep dan/atau aplikasi formulasi.
3	Level 3	T3	Pembuktian konsep fungsi dan/atau karakteristik penting secara analisa dan eksperimental
4	Level 4	T4	Validasi komponen atau subsistem dalam lingkup laboratorium
5	Level 5	T5	Validasi komponen atau subsistem dalam lingkungan yang relevan
6	Level 6	T6	Demonstrasi model atau prototype sistem atau subsistem dalam lingkungan yang relevan
7	Level 7	T7	Demonstrasi model dalam lingkungan yang sebenarnya
8	Level 8	T8	Sistem telah lengkap dan handal melalui pengujian dan demonstrasi dalam lingkungan sebenarnya

9	Level 9	T9	Sistem benar-benar teruji melalui keberhasilan pengoperasian
---	---------	----	--

### 3.1.8 Ground Truth Pengujian Sistem

#### 3.1.8.1 Peringkat QS World University Rankings

Dalam Tugas Akhir ini pemeringkatan keseluruhan dengan menggunakan 6 indikator penilaian mengacu pada kebenaran peringkat QS World University Rankings tahun 2016-2017 sesuai dengan waktu pengerjaan Tugas Akhir ini dimulai. Selain itu melakukan uji coba perbandingan jika mengacu pada Ground Truth peringkat QS World University Rankings dari beberapa tahun yang berbeda. Berikut adalah beberapa peringkat QS World University Rankings dan peringkat reputasi akademik dari berbagai jangka waktu.

##### 1. QS 2015-2015

Peringkat universitas wilayah Indonesia berdasarkan QS World University Rankings 2015-2016.

**Tabel 1-9 Peringkat QS World University Rankings 2015-2016**

<b>Peringkat QS</b>	<b>Nama Perguruan Tinggi</b>
1	Universitas Indonesia (UI)
2	Institut Teknologi Bandung (ITB)
3	Universitas Gadjah Mada (UGM)
4	Universitas Airlangga (UNAIR)
5	Institu Pertanian Bogor (IPB)
6	Universitas Padjadjaran (UNPAD)
7	Universitas Diponegoro (UNDIP)
8	Universitas Muhammadiyah Surakarta (UMS)
9	Institut Teknologi Sepuluh Nopember (ITS)
10	Universitas Brawijaya (UB)

11	Universitas Bina Nusantara (BINUS)
----	------------------------------------

**Tabel 1-10 Peringkat Reputasi Akademik QS 2015-2016**

<b>Peringkat Reputasi Akademik</b>	<b>Nama Perguruan Tinggi</b>
1	Institut Teknologi Bandung (ITB)
2	Universitas Indonesia (UI)
3	Universitas Gadjah Mada (UGM)
4	Universitas Muhammadiyah Surakarta (UMS)
5	Universitas Airlangga (UNAIR)
6	Institu Pertanian Bogor (IPB)
7	Universitas Diponegoro (UNDIP)
8	Universitas Brawijaya (UB)
9	Institut Teknologi Sepuluh Nopember (ITS)
10	Universitas Padjadjaran (UNPAD)
11	Universitas Islam Indonesia (UII)
12	Universitas 17 Agustus 1945 (UNTAG)
13	Universitas Bina Nusantara (BINUS)
14	Universitas Pelita Harapan (UPH)
15	Universitas Sebelas Maret (UNS)
16	Universitas Mataram (UNRAM)

## 2. QS 2016-2017

Peringkat universitas wilayah Indonesia berdasarkan QS World University Rankings 2016-2017.

**Tabel 1-11 Peringkat QS World University Rankings 2016-2017**

<b>Peringkat QS</b>	<b>Nama Perguruan Tinggi</b>
1	Universitas Indonesia (UI)
2	Institut Teknologi Bandung (ITB)

3	Universitas Gadjah Mada (UGM)
4	Universitas Airlangga (UNAIR)
5	Institu Pertanian Bogor (IPB)
6	Universitas Diponegoro (UNDIP)
7	Institut Teknologi Sepuluh Nopember (ITS)
8	Universitas Muhammadiyah Surakarta (UMS)
9	Universitas Brawijaya (UB)

**Tabel 1-12 Peringkat Reputasi Akademik QS 2016-2017**

<b>Peringkat Reputasi Akademik</b>	<b>Nama Perguruan Tinggi</b>
1	Institut Teknologi Bandung (ITB)
2	Universitas Indonesia (UI)
3	Universitas Gadjah Mada (UGM)
4	Universitas Airlangga (UNAIR)
5	Institu Pertanian Bogor (IPB)
6	Universitas Diponegoro (UNDIP)
7	Institut Teknologi Sepuluh Nopember (ITS)
8	Universitas Muhammadiyah Surakarta (UMS)
9	Universitas Brawijaya (UB)

3. QS 2017-2018

Peringkat universitas wilayah Indonesia berdasarkan QS World University Rankings 2017-2018.

**Tabel 1-13 Peringkat QS World University Rankings 2017-2018**

<b>Peringkat QS</b>	<b>Nama Perguruan Tinggi</b>
1	Universitas Indonesia (UI)
2	Institut Teknologi Bandung (ITB)

3	Universitas Gadjah Mada (UGM)
4	Universitas Airlangga (UNAIR)
5	Institu Pertanian Bogor (IPB)
6	Universitas Diponegoro (UNDIP)
7	Institut Teknologi Sepuluh Nopember (ITS)
8	Universitas Muhammadiyah Surakarta (UMS)
9	Universitas Brawijaya (UB)

**Tabel 1-14 Peringkat Reputasi Akademik QS 2017-2018**

<b>Peringkat Reputasi Akademik</b>	<b>Nama Perguruan Tinggi</b>
1	Institut Teknologi Bandung (ITB)
2	Universitas Gadjah Mada (UGM)
3	Universitas Indonesia (UI)
4	Universitas Airlangga (UNAIR)
5	Institu Pertanian Bogor (IPB)
6	Universitas Diponegoro (UNDIP)
7	Institut Teknologi Sepuluh Nopember (ITS)
8	Universitas Muhammadiyah Surakarta (UMS)
9	Universitas Brawijaya (UB)

### 3.1.8.2 Analisa Label Topik Dokumen

Penggunaan teks mining menggunakan metode Probabilistic Latent Semantic Analysis (pLSA) membutuhkan acuan kebenaran sebagai perbandingan apakah keluaran sistem yang dibangun sudah sesuai dengan kebenaran dalam melakukan penentuan topik dalam indikator reputasi akademik.

Ground Truth yang digunakan adalah hasil analisa manual terhadap dokumen abstrak *paper* akademisi perguruan tinggi.

Terdapat 800 dokumen sebagai dataset metode teks mining dari masing-masing perguruan tinggi, dan dianalisa manual untuk mencari label topik apa yang sesuai. Berikut adalah beberapa contoh hasil analisa manual terhadap dataset.

**Tabel 1-15 Contoh Hasil Analisa Manual Sebagai Ground Truth Reputasi Akademik**

No	Nama Perguruan Tinggi	No Dok	Nama Dokumen	Hasil Analisa
1	Institut Pertanian Bogor	1	IPB_001_01	T1, T2
		2	IPB_001_02	T1, T2
		3	IPB_001_03	T1, T2, T3
		4	IPB_05_10	T3, T4, T5
		5	IPB_05_14	T4, T5, T6, T7
2	Institut Teknologi Bandung	1	ITB_01_01	T4, T3, T5
		2	ITB_01_05	T4, T5
		3	ITB_01_22	T1, T2, T3
		4	ITB_01_09	T4, T5, T6, T7
		5	ITB_01_07	T4, T5

### 3.1.9 Dataset Pengujian Sistem

#### 3.1.9.1 Data Dokumen Abstrak Akademisi Perguruan Tinggi

Dokumen abstrak paper akademisi didapatkan dari database metadata publikasi Google Scholar. Dalam Tugas Akhir ini memfokuskan dokumen abstrak paper pada akademisi yang mempunyai bidang fokus teknik. Masing-masing perguruan tinggi diambil 50 dokumen abstrak *paper* teknik berdasarkan sitasi tertinggi dari akademisi yang terdaftar atas institusi perguruan tinggi di Google Scholar.

Dalam proses teks mining, 50 buah dataset abstrak paper perguruan tinggi digabung menjadi satu dokumen utuh yang memungkinkan adanya kata yang sama berjumlah lebih dari satu (duplikasi). Satu dokumen utuh ini selanjutnya akan menjadi

masuk untuk proses tek mining tahap preprocessing. Penentuan dataset dokumen abstrak paper akademisi mempunyai rule seperti berikut:

1. Pencarian daftar akademisi dalam Google Scholar dengan keyword email yang terhubung dengan institusi atau keyword hosting institusi. Misal Institut Teknologi Sepuluh Nopember dicari menggunakan keyword “@its.ac.id” atau www.its.ac.id
2. Daftar akademisi yang diambil adalah yang mempunyai tingkat sitasi tertinggi dan berfokus pada bidang teknik.
3. Masing-masing dari akademisi yang mempunyai sitasi tertinggi diambil paper yang membahas teknik.
4. Dokumen abstrak paper difokuskan pada dokumen berbahasa Inggris.

Tahap proses teks mining dalam Tugas Akhir ini akan menghasilkan reduksi terhadap kata (*term*) yang ada di dalam dokumen abstrak *paper* akademisi. Reduksi terhadap kata (*term*) berperan penting untuk menyeleksi kata (*term*) yang tidak penting. Dilakukan analisa terhadap keberadaan kata (*term*) yang penting dan tidak penting dalam dataset untuk melihat rata-rata duplikasi term. Rincian jumlah kata pada dataset abstrak paper akademisi (16 perguruan tinggi teratas berdasarkan *Ground Truth*) seperti pada Tabel 1-16.

**Tabel 1-16 Daftar Rincian Jumlah Kata Dataset Abstrak Paper Perguruan Tinggi**

No.	Perguruan Tinggi	Total Kata
1.	Institus Pertanian Bogor	10539
2.	Institut Teknologi Bandung	9287
3.	Institut Teknologi Sepuluh Nopember	9630
4.	Universitas 17 Agustus 1945	9146
5.	Universitas Airlangga	9457
6.	Universitas Bina Nusantara	7236
7.	Universitas Brawijaya	8651
8.	Universitas Diponegoro	8397
9.	Universitas Gadjah Mada	10214

10.	Universitas Indonesia	8823
11.	Universitas Islam Indonesia	7897
12.	Universitas Mataram	7527
13.	Universitas Muhammadiyah Surakarta	10330
14.	Universitas Padjadjaran	7774
15.	Universitas Pelita Harapan	7204
16.	Universitas Sebelas Maret	7807

### 3.1.9.2 Data Corpus Taxonomy Bloom

Corpus Taxonomy Bloom berfungsi sebagai dasar pengetahuan corpus tingkat kesiapterapan teknologi (TKT). Corpus Taxonomy Bloom mempunyai 6 level tingkatan yaitu Knowledge, Comprehension, Application, Analysis, Synthesis, dan Evaluation [29]. 6 level corpus Taxonomy Bloom akan digunakan untuk pengetahuan corpus TKT yang nantinya akan dikembangkan lagi menjadi 9 level (*class*). Dilihat dari jumlah keyword yang tercakup dalam masing-masing level Taxonomy Bloom dapat dikatakan kurang banyak dalam jumlah keyword jika dibandingkan dengan jumlah kata (*term*) pada dokumen dataset abstrak *paper* akademisi. Untuk selanjutnya akan diadakan proses breakdown corpus Taxonomy Bloom menjadi corpus TKT dengan melalui berbagai tahap. Rincian jumlah dari masing-masing corpus Taxonomy Bloom seperti pada Tabel 1-17.

**Tabel 1-17 Daftar Rincian Jumlah Kata Corpus Taxonomy Bloom**

No.	Level Taxonomy	Total Kata
1.	Knowledge	35
2.	Comprehension	29
3.	Application	36
4.	Analysis	51
5.	Synthesis	51
6.	Evaluation	46



### 3.1.9.3 Data Corpus Tingkat Kesiapterapan Teknologi (TKT)

Corpus tingkat kesiapterapan teknologi (TKT) merupakan corpus utama untuk proses pengenalan kata (*term*) dalam dokumen abstrak *paper* akademisi agar masuk dalam kategori level TKT yang mana. Corpus TKT ini berjumlah 9 level, dan didapatkan dari proses breakdown 6 level corpus Taxonomy Bloom.

Metode breakdown yang digunakan dalam penentuan corpus TKT ini adalah sebagai berikut:

1. Menggabungkan 6 level corpus Taxonomy Bloom menjadi satu dokumen besar.
2. Mengurutkan kata (*term*) dari corpus Taxonomy Bloom besar berdasarkan bobot konteks kata (*term*) dari yang paling dasar hingga tingkat penerapan. Tahap ini menggunakan kemampuan manusia (*expert*) untuk membedakan dan mensorting sesuai urutan.
3. Menghapus kata yang berjumlah lebih dari satu (duplikasi).
4. Memecah urutan konten kata (*term*) tersebut menjadi 9 bagian. Pemecahan ini didasarkan pada setiap urutan level TKT akan bergantung pada TKT selanjutnya. Sebagai contoh corpus TKT 2 pasti berkorelasi dengan TKT 1 dan TKT 3. Sehingga pemecahan ini mudah untuk dilakukan. Secara tidak langsung level TKT dapat diperoleh kata-kata dari berbagai macam kata pada level Taxonomy Bloom.
5. Dokumen hasil breakdown berjumlah 9 corpus kata TKT baru tanpa mengandung kata duplikasi.

Perbandingan dari 6 level corpus Taxonomy Bloom dengan 9 level corpus TKT seperti pada Tabel 1-18.

**Tabel 1-18 Perbandingan Jumlah Kata Corpus TKT**

Level Taxonomy Bloom	Jumlah Kata	Level TKT	Jumlah Kata
Knowledge	35	TKT 1	31
		TKT 2	24
Comprehension	29	TKT 3	32

Application	36	TKT 4	15
		TKT 5	32
Analysis	51	TKT 6	24
Synthesis	51	TKT 7	26
		TKT 8	30
Evaluation	46	TKT 9	34
Total	256	Total	256

## 3.2 Tahap Perancangan

Tahap ini meliputi perancangan basis data, perancangan tampilan antarmuka (*Interfaces*), perancangan alur proses penggunaan sistem antar pengguna, dan proses utama dalam proses teks mining topik modelling menggunakan metode PLSA. Diharapkan dapat memenuhi tujuan dari pengembangan aplikasi ini.

### 3.2.1 Perancangan Basis Data

Perangkat lunak yang dibangun menggunakan basis data MySQL dengan bahasa syntax SQL. Terdapat 28 tabel untuk menampung hasil proses grabbing data online, hasil proses teks minning berupa skor reputasi akademik, hasil penampung skor masing-masing kriteria, serta data perguruan tinggi. Terdapat 2 buah kegunaan utama dari basis data yang digunakan dalam pembuatan perangkat lunak, yaitu sebagai penyimpan data final, dan penyimpan data temporary update. Dengan rincian 16 tabel utama dan 12 tabel temporary sebagai penampung sementara data hasil Teknik grabbing.

#### 3.2.1.1 Tabel Universitas

Identifikasi : universitas

Deskripsi isi : tabel untuk menyimpan data nama perguruan tinggi beserta informasi detail

Id_field	Deskripsi	Tipe & length	Null
----------	-----------	---------------	------

Id	Variabel primary key	Int(2)	No
Nama	Variabel penampung nama	Varchar(255)	No
Domain	Variabel penampung domain webhost	Varchar(150)	No
url_web	Variabel penampung alamat link webhost	Varchar(150)	No
Kota	Variabel kota	Varchar(150)	No
provinsi	Variabel provinsi	Varchar(150)	No

### 3.2.1.2 Tabel Reputasi Akademik

Identifikasi : reputasi\_akademik

Deskripsi isi : tabel untuk menyimpan nilai hasil proses teks mining metode PLSA sebagai representasi nilai reputasi akademik

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel berisi nama perguruan tinggi	Varchar(255)	No
Skor_universitas	Variabel penampung skor final reputasi akademik	Float	No
Bobot_total	Variabel penampung jumlah hasil reputasi akademik	Int(2)	No
Jumlah_dokumen	Variabel penampung jumlah dokumen olah	Int(2)	No
Tanggal_akses	Variabel timestamp	Datetime	No

### 3.2.1.3 Tabel Akreditasi

Identifikasi : tampung\_akreditasi

Deskripsi isi : tabel untuk menyimpan nilai hasil proses grabbing akreditasi perguruan tinggi

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel berisi nama perguruan tinggi	Varchar(255)	NO
Skor_universitas	Variabel penampung skor final reputasi akademik	Float	NO

Bobot_total	Variabel penampung jumlah hasil reputasi akademik	Int(2)	No
Jumlah_dokumen	Variabel penampung jumlah dokumen olah	Int(2)	No
Tanggal_akses	Varabel waktu akses	Datetime	No

#### 3.2.1.4 Tabel Jumlah Mahasiswa

Identifikasi : jumlah\_mahasiswa

Deskripsi isi : tabel untuk menyimpan jumlah mahasiswa yang didapatkan dari data PDDIKTI

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Kode_universitas	Variabel penampung kode status diakui universitas	Int(2)	No
Nama_universitas	Variabel berisi nama perguruan tinggi	Varchar(255)	No
Jumlah_tahun1_semester1	Variabel jumlah mahasiswa tahun1 pada semester1	Int(2)	No
Jumlah_tahun1_semester2	Variabel jumlah mahasiswa tahun1 pada semester2	Int(2)	No
Jumlah_tahun2_semester1	Variabel jumlah mahasiswa tahun2 pada semester1	Int(2)	No
Jumlah_tahun2_semester2	Variabel jumlah mahasiswa tahun2 pada semester2	Int(2)	No
Status_universitas	Variabel berisi keterangan status universitas	Varchar(5)	No

#### 3.2.1.5 Tabel Jumlah Sitasi Sinta

Identifikasi : jumlah\_sitasi\_sinta

Deskripsi isi : tabel untuk menyimpan jumlah sitasi akademisi yang didapatkan dari database online SINTA

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Kode_universitas	Variabel penampung kode status diakui universitas	Int(2)	No
Nama_universitas	Variabel berisi nama perguruan tinggi	Varchar(250)	No
Jumlah_tahun1_semester1	Variabel jumlah mahasiswa tahun1 pada semester1	Int(2)	No
Jumlah_tahun1_semester2	Variabel jumlah mahasiswa tahun1 pada semester2	Int(2)	No
Jumlah_tahun2_semester1	Variabel jumlah mahasiswa tahun2 pada semester1	Int(2)	No
Jumlah_tahun2_semester2	Variabel jumlah mahasiswa tahun2 pada semester2	Int(2)	No
Status_universitas	Variabel berisi keterangan status universitas	Varchar(5)	No

### 3.2.1.6 Tabel Jumlah Author Sinta

Identifikasi : jumlah\_author\_sinta

Deskripsi isi : tabel untuk menyimpan jumlah h-index dan i10-index akademisi yang didapatkan dari database online SINTA

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Id_data	Variabel tampung id data secara kolektif	Int(2)	No
Nomor	Variabel nomor akademisi dalam range satu kali iterasi	Int(2)	No
Nama_author	Variabel penampung nama akademisi	Varchar(250)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(250)	No
Jumlah_sitasi_scholar	Variabel jumlah sitasi	Int(2)	No

I10_index_scholar	Vaiabel jumlah i10-index	Int(2)	No
H_index_scholar	Variabel jumlah h_index	Int(2)	No
Waktu_akses	Variabel penampung waktu akses	Tinytext	No

### 3.2.1.7 Tabel Precense

Identifikasi : precense

Deskripsi isi : tabel untuk menyimpan jumlah domain, subdomain dan repository webhost perguruan tinggi yang terindex Google

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
url_web	Variabel penampung link webhost	Varchar(255)	No
Domain_search	Variabel penampung link search webhost	Varchar(150)	No
Hasil_search	Variabel penampung hasil search webhost	Int(2)	No
Tanggal_akses	Variabel waktu akses	Datetime	No

### 3.2.1.8 Tabel Impact

Identifikasi : impact

Deskripsi isi : tabel untuk menyimpan jumlah page authority, domain authority dan backlink webhost perguruan tinggi

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
url_web	Variabel penampung link webhost	Varchar(255)	No
Domain_web	Variabel penampung link webhost tanpa “www”	Varchar(150)	No
Page_authority	Variabel penampung nilai page authority	Int(2)	No

Domain_authority	Variabel penampung nilai domain authority	Int(2)	No
Backlink_eksterna	Variabel penampung jumlah backlink	Int(2)	No
Tanggal_aks	Variabel waktu akses	Datetime	No

### 3.2.1.9 Tabel Alexa Rank

Identifikasi : alexa\_rank

Deskripsi isi : tabel untuk menyimpan peringkat webhost berdasarkan alexa rank

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
url_web	Variabel penampung link webhost	Varchar(255)	No
Rank_world	Variabel penampung peringkat webhost skala dunia	Int(2)	No
Rank_country	Variabel penampung peringkat webhost skala nasional	Int(2)	No
Country_name	Variabel penampung nilai domain authority	Varchar(150)	No
Tanggal_aks	Variabel waktu akses	Datetime	No

### 3.2.1.10 Tabel Kumpulan Skor Final

Identifikasi : kumpulan\_skor\_final

Deskripsi isi : tabel untuk menyimpan skor dari setiap kriteria pemeringkatan

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Kode_universitas	Variabel penampung kode status diakui universitas	Float	No
Nama_universitas	Variabel nama perguruan tinggi	Varchar(255)	No

Skor_reputasi_akademik_normalisasi	Variabel penampung skor reputasi akademik yang dinormalisasi	Float	No
Skor_nilai_akreditasi_normalisasi	Variabel penampung skor nilai akreditasi yang dinormalisasi	Float	No
Skor_jumlah_mahasiswa_normalisasi	Variabel penampung skor jumlah mahasiswa yang dinormalisasi	Float	No
Skor_jumlah_sitasi_normalisasi	Variabel penampung skor jumlah sitasi yang dinormalisasi	Float	No
Skor_jumlah_tenaga_pengajar_normalisasi	Variabel penampung skor jumlah tenaga pengajar yang dinormalisasi	Float	No
Skor_kualitas_sitasi_normalisasi	Variabel penampung skor kualitas sitasi yang sudah dinormalisasi	Float	No
Skor_jumlah_phd_normalisasi	Variabel penampung skor jumlah tenaga pengajar S3 (Phd) yang sudah dinormalisasi	Float	No
Skor_jumlah_paper_normalisasi	Variabel penampung skor jumlah paper yang dinormalisasi	Float	No
Skor_jumlah_jurnal_artikel	Variabel penampung skor jumlah jurnal artikel yang dinormalisasi	Float	No
Skor_jumlah_paper_conference	Variabel penampung skor jumlah paper conference yang dinormalisasi	Float	No
Skor_precense_normalisasi	Variabel penampung skor precense yang dinormalisasi	Float	No
Skor_impact_normalisasi	Variabel penampung skor impact yang dinormalisasi	Float	No



Skor_transpa rency_norma lisasi	Variabel penampung skor transparency yang dinormalisasi	Float	No
Skor_exellec en_normalis asi	Variabel penampung skor excellence yang dinormalisasi	Float	No
Skor_moz_a nalytic_nor malisasi	Variabel penampung skor moz analytic yang dinormalisasi	Float	No
Skor_alexa_ rank_normal isasi	Variabel penampung skor alexa rank yang dinormalisasi	Float	No
Skor_kualita s_dosen_nor malisasi	Variabel penampung skor kualitas dosen yang dinormalisasi	Float	No
Skor_kecuku pan_dosesn_ normalisasi	Variabel penampung skor kecukupan dosesn yang dinormalisasi	Float	No
Skor_kualita s_manajeme n_normalisa si	Variabel penampung skor kualitas manajemen yang dinormalisasi	Float	No
Skor_kualita s_kegiatan_k emahasiswaa n	Variabel penampung skor kualitas kegiatan mahasiswa yang dinormalisasi	Float	No
Sko_kualitas _kegiatan_p enelitian	Variabel penampung skor kualitas kegiatan penelitian yang dinormalisasi	Float	No

### 3.2.1.11 Tabel Peringkat 4ICU

Identifikasi : 4icu

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan 4ICU

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_unive rsitas	Variabel penampung nama universitas	Varchar(255)	No

Skor_4icu	Variabel penampung skor 4ICU	Float	No
Peringkat_4icu	Variabel penampung peringkat 4ICU	Int(2)	No

### 3.2.1.12 Tabel Peringkat DIKTI

Identifikasi : dikti

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan DIKTI Indonesia

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(150)	No
Skor_dikti	Variabel penampung skor DIKTI	Float	No
Peringkat_dikti	Variabel penampung peringkat DIKTI	Int(2)	No

### 3.2.1.13 Tabel Peringkat QS General

Identifikasi : qs\_general

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan QS rank versi general

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(255)	No
Skor_qs	Variabel penampung skor QS General	Float	No
Peringkat_qs	Variabel penampung peringkat QS General	Int(2)	No

### 3.2.1.14 Tabel Peringkat QS Asia

Identifikasi : qs\_asia

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan QS rank versi region Asia

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(255)	No
Skor_qsasia	Variabel penampung skor QS Asia	Float	No
Peringkat_qs-asia	Variabel penampung peringkat QS Asia	Int(2)	No

### 3.2.1.15 Tabel Peringkat QS Metode PLSA

Identifikasi : qs\_plsa

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan QS rank berdasarkan metode PLSA

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(255)	No
Skor_plsa	Variabel penampung skor QS PLSA	Float	No
Peringkat_plsa	Variabel penampung peringkat QS PLSA	Int(2)	No

### 3.2.1.16 Tabel Peringkat Webometrics

Identifikasi : webometrics

Deskripsi isi : tabel untuk menyimpan skor dan peringkat pemeringkatan Webometrics

Id_field	Deskripsi	Tipe & length	Null
Id	Variabel primary key	Int(2)	No
Nama_universitas	Variabel penampung nama universitas	Varchar(255)	No
Skor_webometrics	Variabel penampung skor Webometrics	Float	No
Peringkat_webometrics	Variabel penampung peringkat Webometrics	Int(2)	No

### 3.2.2 Perancangan Antarmuka

Perangkat lunak yang dibangun berbasis pada aplikasi website. Antarmuka perangkat lunak dibangun menggunakan bahasa pemrograman PHP dan HTML. Berikut adalah rincian antarmuka yang dibangun dalam perangkat lunak.

#### 3.2.2.1 Antarmuka Home

Identifikasi : antarmuka home (index)

Deskripsi isi : antarmuka ini berfungsi sebagai halaman root utama dalam perangkat lunak website. Terdapat satu kebutuhan fungsional dalam antarmuka ini yaitu pencarian perguruan tinggi berdasarkan potongan nama. Terdapat kolom search perguruan tinggi.

#### 3.2.2.2 Antarmuka Latar Belakang Pemeringkatan

Identifikasi : antarmuka sekilas pemeringkatan

Deskripsi isi : antarmuka ini berfungsi sebagai penampil informasi sederhana yang menyatakan latar belakang kegunaan dari sistem pemeringkatan secara general.

#### 3.2.2.3 Antarmuka Metode Penilaian

Identifikasi : antarmuka metode penilaian

Deskripsi isi : antarmuka ini berfungsi sebagai penampil informasi mengenai kriteria penilaian apa saja yang digunakan dalam sistem pemeringkatan. Terdapat informasi secara general dan informasi secara detail dari setiap tipe pemeringkatan, mulai dari pemeringkatan QS General, QS Asia, Webometrics, 4ICU, pemeringkatan DIKTI, dan QS menggunakan metode PLSA. Terdapat tombol untuk melihat tipe pemeringkatan yang diinginkan.

#### 3.2.2.4 Antarmuka Peringkat Berdasarkan Region

Identifikasi : antarmuka peringkat berdasarkan region

Deskripsi isi : antarmuka ini berfungsi sebagai halaman pencarian peringkat perguruan tinggi berdasarkan region wilayah provinsi perguruan tinggi. Dengan menampilkan desain map negara Indonesia sebagai inputan pencarian yang mudah untuk dipahami. Pencarian dilakukan dengan cara menekan salah satu bagian provinsi dari map negara Indonesia.

#### 3.2.2.5 Antarmuka Hasil Pencarian Berdasar Region Provinsi

Identifikasi : antarmuka hasil pencarian berdasarkan region provinsi

Deskripsi isi : antarmuka ini berfungsi sebagai halaman penampil hasil pencarian perguruan tinggi berdasarkan region provinsi yang dimiliki. Ditampilkan juga peringkat dari perguruan tinggi dalam hasil pencarian dari beberapa tipe pemeringkatan seperti QS General, QS Asia, Webometrics, 4ICU, pemeringkatan DIKTI, dan QS PLSA. Terdapat tombol untuk melihat tipe pemeringkatan yang diinginkan, dan terdapat kolom filter untuk mencari berdasarkan kota.

#### 3.2.2.6 Antarmuka Hasil Pencarian Berdasar Pencarian Keyword Potongan Nama

Identifikasi : antarmuka hasil pencarian berdasarkan pencarian keyword potongan nama

Deskripsi isi : antarmuka ini berfungsi sebagai halaman penampil hasil pencarian perguruan tinggi berdasarkan keyword potongan nama. Terdapat informasi peringkat dari masing-masing perguruan tinggi yang dihasilkan. Terdapat tombol “Lihat Detail” untuk melihat detail informasi akademik perguruan tinggi.

#### 3.2.2.7 Antarmuka Detail Informasi Akademik

Identifikasi : antarmuka detail informasi akademik

Deskripsi isi : antarmuka ini berfungsi sebagai halaman penampil informasi akademik perguruan tinggi. Terdapat beberapa informasi diantaranya mahasiswa, kecukupan dosen, status akreditasi, kualitas akademisi, produktifitas institusi, dan kepopuleran website. Ditampilkan juga grafik peringkat dari berbagai tipe pemeringkatan.

### 3.2.2.8 Antarmuka Pembobotan Pemeringkatan

Identifikasi : antarmuka pembobotan pemeringkatan

Deskripsi isi : antarmuka ini berfungsi sebagai halaman masukan bobot pemeringkatan yang diinginkan. Terdapat tabel tipe pemeringkatan dan mempunyai daftar kriteria yang digunakan. Pengguna dapat memasukkan bobot pemeringkatan sesuai kebutuhan. Nilai maksimal pembobotan dari satu tipe pemeringkatan adalah 100%, jika lebih dari itu maka akan muncul pesan error bobot berlebihan. Terdapat tombol pencarian berdasarkan tipe pemeringkatan, diantaranya “Cek QS-General”, “Cek QS-Asia”, “Cek 4ICU”, “Cek Webometrics”, “Cek DIKTI”, “Cek QS PLSA”, “Cek QS PLSa”, “Cek QS Helmholtz”, “Cek QS Adaboost”.

### 3.2.2.9 Antarmuka Hasil Pemeringkatan Berdasar Pembobotan Tipe Pemeringkatan yang Diinginkan

Identifikasi : antarmuka hasil pemeringkatan berdasarkan hasil pembobotan

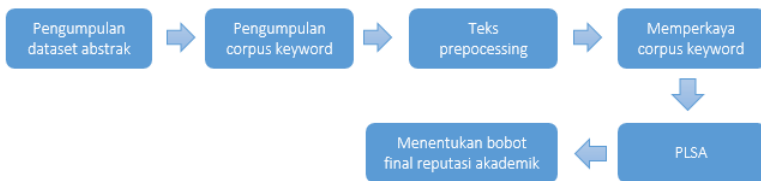
Deskripsi isi : antarmuka ini berfungsi sebagai halaman penampil hasil pembobotan perguruan tinggi sesuai dengan tipe pemeringkatan yang dikehendaki. Terdapat kolom filter untuk mencari berdasarkan kota.

### 3.2.3 Perancangan Kriteria Penilaian Perguruan Tinggi

Dalam menentukan peringkat perguruan tinggi menggunakan metode PLSA, dibutuhkan beberapa kriteria penilaian lain sebagai penunjang aspek kriteria kuantitatif dan kualitatif [4]. Penjelasan masing-masing kriteria akan dijelaskan dibawah ini.

#### 3.2.3.1 Reputasi Akademik

Merupakan kriteria utama yang bersifat kualitatif yang didapatkan dari proses teks mining dokumen abstrak perguruan tinggi menggunakan metode Probabilistic Latent Semantic Analysis (pLSA) [6]. Mempunyai bobot pemeringkatan 40% dari total bobot keseluruhan. Alur proses menghasilkan kriteria reputasi akademik seperti pada Gambar 3.2-1.



**Gambar 3.2-1 Alur Menentukan Reputasi Akademik**

Langkah pertama dengan mengumpulkan dataset berupa abstrak paper perguruan tinggi, kemudian membangun corpus keyword yang didapatkan dari taxonomy Bloom dan dikembangkan menjadi corpus TKT seperti pada langkah 4 pada bagan diatas. Selanjutnya melakukan tahap teks preprocessing untuk menghasilkan kata inputan yang baku dari dokumen dataset. Hasil dari tahap teks preprocessing akan digunakan sebagai penentuan asumsi label topik awal, dimana setiap dokumen harus mempunyai label sebagai acuan topiknya. Kemudian masuk pada tahap inferensi Gibbs Sampling untuk menghasilkan probabilitas terbaik dari dokumen terhadap topik. Hasil dari tahap ini berupa

probabilitas akhir yang nantinya akan dihitung menggunakan pembobotan final reputasi akademik. Hasil ini kemudian disimpan dalam database sebagai nilai indicator reputasi akademik.

### 3.2.3.2 Akreditasi Perguruan Tinggi

Merupakan kriteria yang bersifat kuantitatif. Didapatkan dari proses grabbing data online pada website BAN-PT. Data yang diambil adalah data nama perguruan tinggi, masa berlaku, nilai huruf akreditasi, serta status perguruan tinggi. Mempunyai bobot 10% dari total bobot keseluruhan. Alur proses menghasilkan kriteria akreditasi perguruan tinggi seperti pada Gambar 3.2-2.



**Gambar 3.2-2 Alur Menentukan Skor Akreditasi**

Langkah pertama dengan menyiapkan situs website yang akan dijadikan sasaran *grabbing*, dalam Tugas Akhir ini menggunakan 2 percobaan pada website BAN-PT dan website penyedia informasi akreditasi. Kemudian Menentukan teknik grabbing apa dan library yang akan digunakan, dalam Tugas Akhir ini menggunakan Bahasa pemrograman PHP, library HTML DOM Parser, dan beberapa fungsi bawaan dari PHP sendiri. Kemudian melakukan grabbing dengan ketentuan harus mempunyai bandwidth yang cukup. Karena proses grabbing membutuhkan waktu yang cukup lama melihat konten yang ingin diambil cukup besar dan banyak. Dalam proses grabbing data yang masuk akan langsung disimpan dalam table temporary pada database, sebagai penampung sementara karena tujuan utama disini adalah mengambil sebisa mungkin. Dalam database table temporary ini kemudian akan dianalisa dan dilakukan normalisasi hasil untuk menghasilkan skor akreditasi yang paling akhir. Skor akreditasi ini kemudian akan disimpan lagi dalam database dengan table yang



berbeda sebagai implementasi skor final akreditasi perguruan tinggi.

### 3.2.3.3 Jumlah Rasio Mahasiswa dan Tenaga Pengajar Aktif

Merupakan kriteria yang bersifat kuantitatif. Didapatkan dari data rekap database online PDDIKTI. Data yang diambil merupakan rekap yang berisi jumlah mahasiswa, nama perguruan tinggi serta jumlah tenaga pengajar keseluruhan. Mempunyai bobot 20% dari total bobot keseluruhan. Alur proses menghasilkan kriteria rasio mahasiswa dan tenaga pengajar aktif seperti pada Gambar 3.2-3.



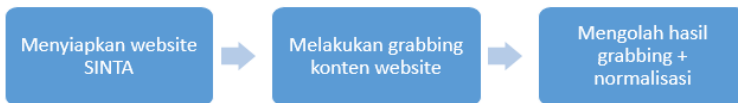
**Gambar 3.2-3 Alur Menentukan Skor Rasio Mahasiswa dan Tenaga Penagajar Aktif**

Langkah pertama dengan mengumpulkan data rakap jumlah mahasiswa dan rekap jumlah tenaga pengajar dari seluruh perguruan tinggi di Indonesia. Data ini disediakan oleh PDDIKTI sebagai pusat databasenya. Untuk mempermudah proses analisa, dalam Tugas Akhir ini menggunakan data rekap berformat .xlsx kemudian akan dibaca menggunakan library excel dalam Bahasa pemrograman python. Karena file excel berupa row dan column maka akan diambil beberapa variable penting sebagai acuan penilaian. Variable jumlah mahasiswa semester 1 tahun 1, mahasiswa semester 2 tahun 1, mahasiswa semester 1 tahun 2, dan semester 2 tahun 2 menjadi acuan panilaian mahasiswa. Serta data jumlah teanga pengajar secara keseluruhan dari setiap perguruan tinggi akan menjdi acuan penilaian jumlah tenaga pangajar. Nilai dari variable ini akan dinormalisasi terlebih dahulu dengan membagi dengan nilai tertinggi. Kemudian akan dirasiokan dengan jumlah tenaga pengajar, kemudian disimpan dalam database

sebagai implementasi skor final rasio mahasiswa dan tenaga pengajar aktif.

#### 3.2.3.4 Jumlah Sitasi Akademisi Penelitian

Merupakan kriteria yang bersifat kuantitatif. Didapatkan dari proses grabbing data online Science and Technology Index (SINTA). Data yang digunakan adalah nama perguruan tinggi, serta jumlah sitasi dan author akademisi yang terindex SINTA. Mempunyai bobot 20% dari total bobot keseluruhan. Alur proses menghasilkan kriteria jumlah sitasi akademisi penelitian seperti pada Gambar 3.2-4.



**Gambar 3.2-4 Alur Menentukan Skor Jumlah Sitasi Akademisi Penelitian**

Langkah pertama dengan menyiapkan website sasaran grabbing yaitu situs website Sciene and Texchnology Index (SINTA) dari Kemenristek Dikti. Dalam situs ini terdapat beberapa variable penting sebagai acuan penilaian indikator. Setiap variable akan dilakukan teknik *grabbing* dan sama seperti pada proses menentukan indikator akreditasi perguruan tinggi. Dalam proses ini juga membutuhkan database table temporary sebagai penampung sementara. Variable yang digunakan adalah jumlah sitasi dari google scholar dan scopus. Kemudian dari table temporary ini akan dianalisa dan dinormalisasi dan dicari rata-ratanya sebagai skor final jumlah sitasi akademisi penelitian dan disimpan lagi dalam table final.

#### 3.2.3.5 Jumlah Tenaga Pengajar (Dosen)

Merupakan kriteria yang bersifat kuantitatif yang didapatkan dari data rekap database online PDDIKTI. Data yang

diambil merupakan jumlah tenaga pengajar dengan berbagai status ketenaga kerjaan, seperti pegawai tetap maupun tidak tetap. Mempunyai bobot 5% dari total bobot keseluruhan. Alur proses menghasilkan kriteria jumlah tenaga pengajar (dosen) seperti pada Gambar 3.2-3 diatas.

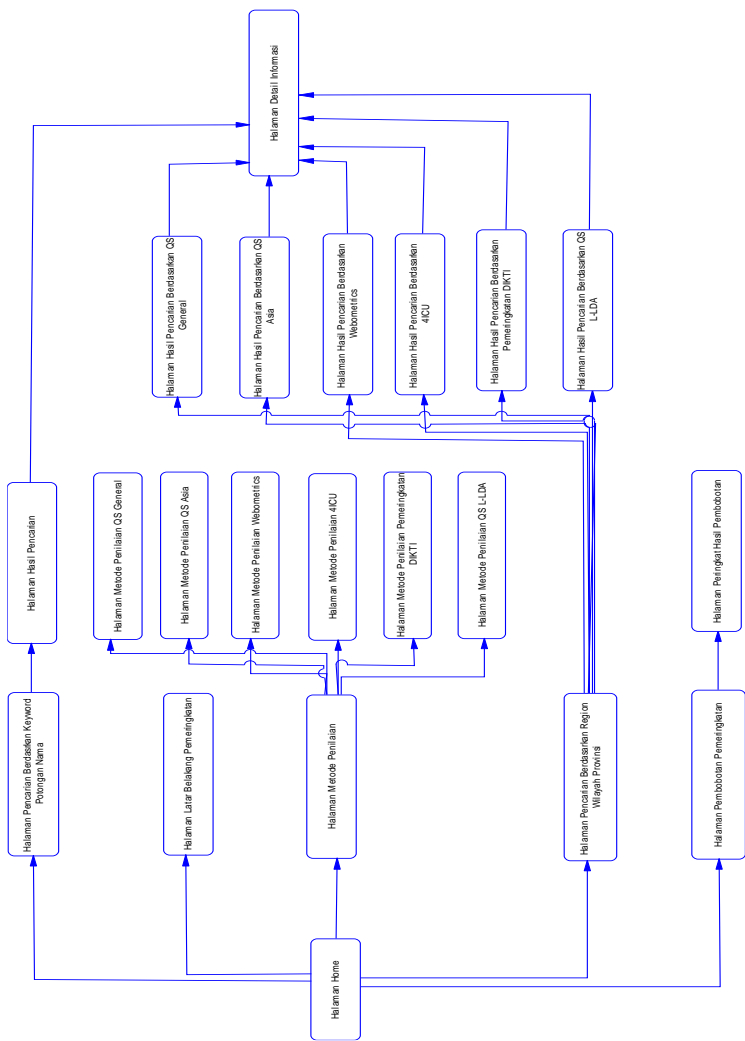
Prinsip kerja Menentukan indikator ini sama seperti indikator jumlah rasio mahasiswa dan tenaga pengajar aktif. Hanya saja disini ada beberapa data rekap yang digunakan sebagai pendukung yaitu rekap jumlah tenaga pengajar yang bersifat tetap dan tidak tetap (honorer). Data ini akan dibaca menggunakan library excel pada Bahasa pemrograman PHP kemudian akan dianalisa dan dinormalisasi untuk menghasilkan jangkauan nilai yang rata. Terdapat variable tanga pengajar tetap dan tidak tetap yang nilai masing-masing telah dinormalisasi, kemudian dijumlahkan dan dirata-rata. Nilai rata-rata ini akan menjadi skor indikator jumlah tenaga pengajar (dosen).

#### 3.2.3.6 Kualitas Sitasi

Merupakan kriteria yang bersifat kuantitatif yang didapatkan dari proses grabbing data online Science and Technology Index (SINTA). Data yang diambil merupakan nilai h-index dan i10-index Google Scholar dari akademisi perguruan tinggi. Mempunyai bobot 5% dari total bobot keseluruhan. Alur proses menghasilkan kriteria kualitas sitasi seperti pada Gambar 3.2-4 diatas.

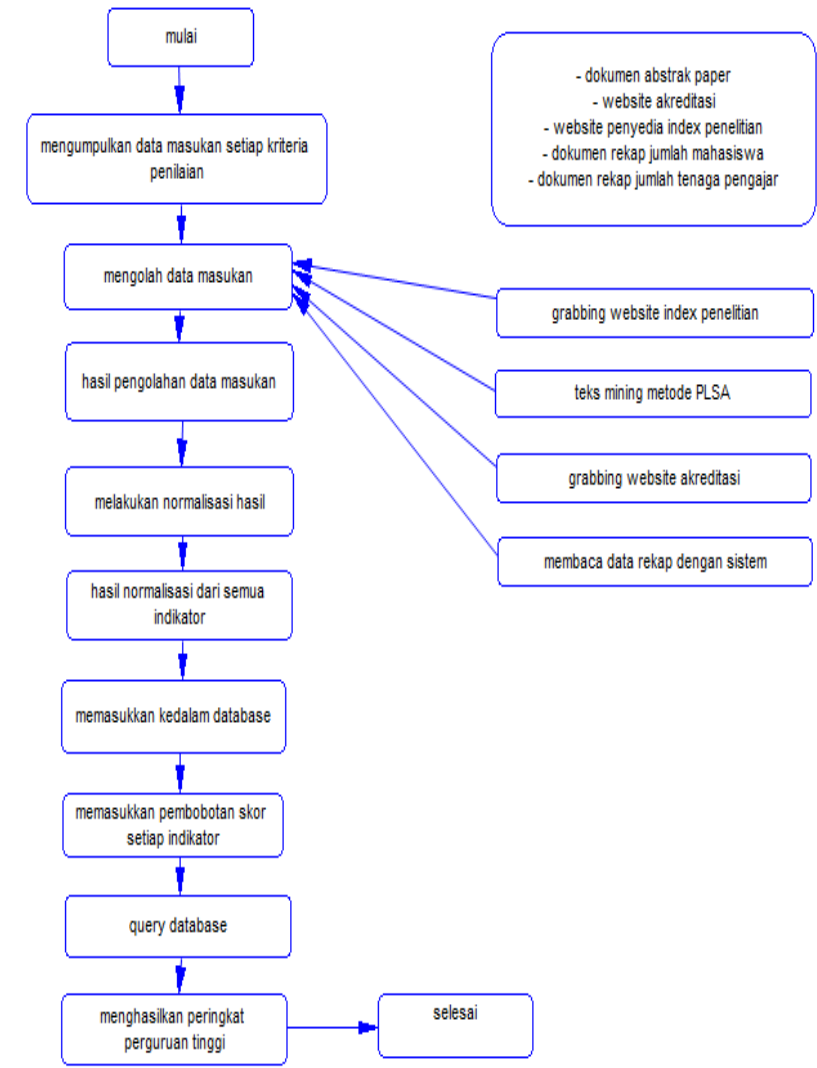
Prinsip kerja dalam Menentukan indikator ini sama seperti pada indikator jumlah sitasi akademisi penelitian. Hanya saja disini yang diambil adalah variable jumlah h-index dan i10-index dari google scholar. Kedua variable ini akan dilakukan normalisasi nilai dan dicari rata-ratanya. Dan kemudian disimpan dalam table database baru sebagai implementasi skor indikator kualitas sitasi.

3.2.4 Proses Sistem Aplikasi Pengguna (Diagram Aplikasi)



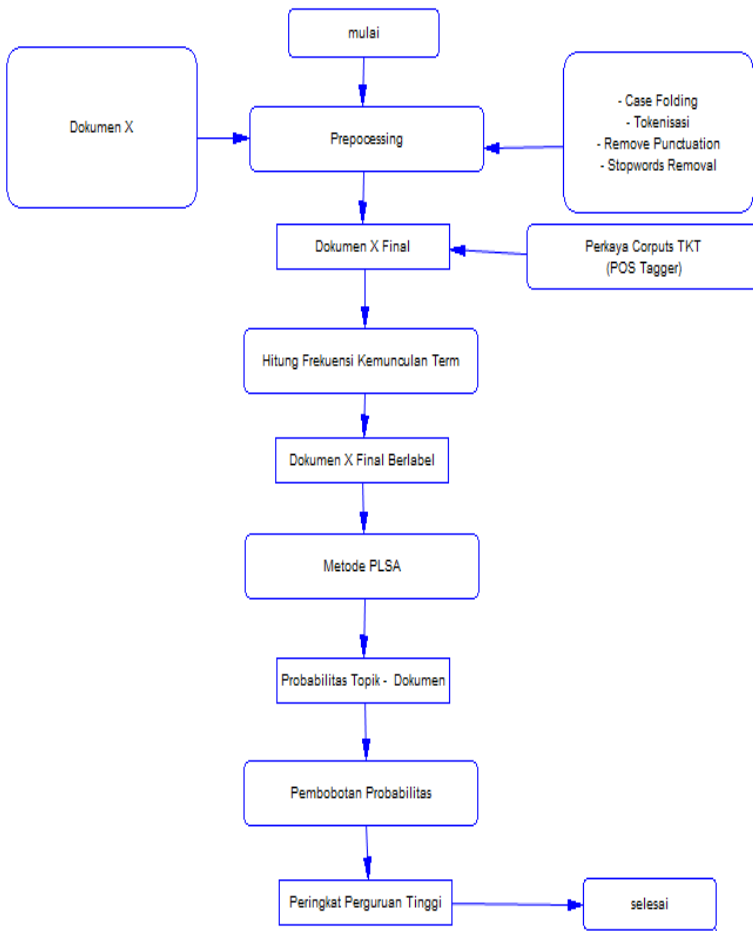
Gambar 3.2-5 Diagram Aplikasi

### 3.2.5 Proses Kerja Sistem Secara Keseluruhan (Flowchart Sistem)



**Gambar 3.2-6 Flowchart Sistem Keseluruhan**

### 3.2.6 Proses Teks Mining Metode PLSA (Flowchart Metode PLSA)



**Gambar 3.2-7 Flowchart Metode PLSA**

*[Halaman ini sengaja dikosongkan]*

## **BAB IV**

### **IMPLEMENTASI SISTEM**

Pada bab ini diuraikan mengenai implementasi perangkat lunak dari rancangan metode yang telah dibahas pada Bab III meliputi lingkungan implementasi, syntax basis data, dan kode program dalam perangkat lunak. Selain itu implementasi dari setiap proses, parameter masukan, keluaran, dan beberapa keterangan pendukung yang berhubungan dengan perangkat lunak.

#### **4.1 Lingkungan Implementasi**

Lingkungan implementasi perangkat lunak yang digunakan untuk mengembangkan Tugas Akhir ini sebagai berikut:

1. 1 Laptop untuk developer dan running perangkat lunak server menggunakan Intel® Core™ i7-5500U CPU @2.40GHz (4CPUs) , RAM 12GB dengan Sistem Operasi Windows 8.1 Profesional x64 (64-bit).
2. 1 Laptop untuk pengguna End User menggunakan Intel® Core™ i7-5500U CPU @2.40GHz (4CPUs) , RAM 12GB dengan Sistem Operasi Windows 8.1 Profesional x64 (64-bit).
3. Database yang digunakan pada localhost adalah MySQL database dengan IDE MySQLyog Community v11.52 (64-bit) sebagai uji coba localhost.
4. Database yang digunakan pada server online adalah MySQL.
5. Webhosting menggunakan server hosting dari namecheap.com
6. Mozilla Firefox 53.0.3 (32-bit) sebagai antarmuka untuk pengujian perangkat lunak bagi pengguna End User.
7. SublimeText 2 v2.0.2 sebagai IDE text editor dari programmer developer.
8. XAMPP v3.2.1 sebagai webserver sebagai server uji coba localhost dari programmer developer.



9. Python v2.7.12 dan Anaconda 4.2.0 (64-bit) sebagai IDE dan bahasa pemrograman utama untuk pengolahan teks minning.
10. StarUML sebagai tools pembuatan diagram Use Case perangkat lunak.
11. Power Designer sebagai tools pembuatan diagram activity CDM dan PDM perangkat lunak.

## 4.2 Implementasi Syntax Basis Data

### 4.2.1 Create Database “database\_qs\_rank”

```
CREATE DATABASE /*!32312 IF NOT
EXISTS*/`database_qs_2` /*!40100 DEFAULT CHARACTER SET
latin1 */;

USE `database_qs_2`;
```

**Kode Sumber 4.2-1 Kode Sumber Create Database**

### 4.2.2 Create Table “universitas”

```
CREATE TABLE `universitas` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama` varchar(255) NOT NULL,
  `domain` varchar(150) NOT NULL,
  `url_web` varchar(150) NOT NULL,
  `kota` varchar(150) NOT NULL,
  `provinsi` varchar(150) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=410 DEFAULT
CHARSET=latin1;
```

**Kode Sumber 4.2-2 Kode Sumber Create Table universitas**

### 4.2.3 Create Table “reputasi akademik”

```
CREATE TABLE `reputasi_akademik` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_universitas` float NOT NULL,
  `bobot_total` int(2) NOT NULL,
  `jumlah_dokumen` int(2) NOT NULL,
  `tanggal_akses` datetime NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=484 DEFAULT
```

#### 4.2.4 Create Tabel “akreditasi”

```
CREATE TABLE `tampung_akreditasi` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nomor` int(3) NOT NULL,
  `nama_universitas` varchar(255) NOT NULL,
  `peringkat` varchar(150) NOT NULL,
  `masa_berlaku` varchar(150) NOT NULL,
  `status` varchar(150) NOT NULL,
  `skor` int(2) NOT NULL,
  `tmp_nama_universitas` text NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=13095 DEFAULT
CHARSET=latin1;
```

**Kode Sumber 4.2-4 Create Tabel akreditasi**

#### 4.2.5 Create Tabel “jumlah\_mahasiswa”

```
CREATE TABLE `jumlah_mahasiswa` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `kode_universitas` int(4) NOT NULL,
  `nama_universitas` varchar(255) NOT NULL,
  `jumlah_tahun1_semester1` int(2) NOT NULL,
  `jumlah_tahun2_semester1` int(2) NOT NULL,
  `jumlah_tahun2_semester2` int(2) NOT NULL,
  `status_universitas` varchar(50) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=42783 DEFAULT
CHARSET=latin1;
```

**Kode Sumber 4.2-5 Create Table jumlah\_mahasiswa**

#### 4.2.6 Create Tabel “jumlah\_tenaga\_pengajar\_jabatan\_akademik”

```
CREATE TABLE
`jumlah_tenaga_pengajar_jabatan_akademik` (
  `id` int(2) NOT NULL AUTO_INCREMENT,
  `kode_universitas` int(2) NOT NULL,
  `nama_universitas` varchar(100) NOT NULL,
  `asisten_ahli_10000` int(2) NOT NULL,
  `asisten_ahli_15000` int(2) NOT NULL,
  `lektor_20000` int(2) NOT NULL,
  `lektor_30000` int(2) NOT NULL,
  `lektor_kepala_40000` int(2) NOT NULL,
  `lektor_kepala_55000` int(2) NOT NULL,
  `lektor_kepala_70000` int(2) NOT NULL,
  `profesor_85000` int(2) NOT NULL,
  `profesor_105000` int(2) NOT NULL,
  `tanpa_jabatan` int(2) NOT NULL,
  `status_universitas` tinytext NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=16840 DEFAULT
CHARSET=latin1;
```

Kode Sumber 4.2-6 Create Table jumlah\_tenaga\_pengajar\_jabatan\_akademik

#### 4.2.7 Create Tabel “jumlah\_tenaga\_pengajar\_tetap”

```
CREATE TABLE `jumlah_tenaga_pengajar_tetap` (
  `id` int(2) NOT NULL AUTO_INCREMENT,
  `kode_universitas` int(2) NOT NULL,
  `nama_universitas` varchar(100) NOT NULL,
  `D1` int(2) NOT NULL,
  `D2` int(2) NOT NULL,
  `D3` int(2) NOT NULL,
  `D4` int(2) NOT NULL,
  `S1` int(2) NOT NULL,
  `S2` int(2) NOT NULL,
```

#### 4.2.8 Create Tabel “jumlah\_tenaga\_pengajar\_tidak\_tetap”

```
CREATE TABLE `jumlah_tenaga_pengajar_tidak_tetap`
(
  `id` int(2) NOT NULL AUTO_INCREMENT,
  `kode_universitas` int(2) NOT NULL,
  `nama_universitas` varchar(100) NOT NULL,
  `D1` int(2) NOT NULL,
  `D2` int(2) NOT NULL,
  `D3` int(2) NOT NULL,
  `D4` int(2) NOT NULL,
  `S1` int(2) NOT NULL,
  `S2` int(2) NOT NULL,
  `S3` int(2) NOT NULL,
  `non_formal` int(2) NOT NULL,
  `informal` int(2) NOT NULL,
  `lainnya` int(2) DEFAULT NULL,
  `SP1` int(2) NOT NULL,
  `SP2` int(2) NOT NULL,
  `profesi` int(2) NOT NULL,
  `tanpa_jenjang` int(2) NOT NULL,
  `status_universitas` tinytext NOT NULL,
  PRIMARY KEY (`id`)
```

#### 4.2.9 Create Tabel “jumlah\_sitasi\_sinta”

```
CREATE TABLE `jumlah_sitasi_sinta` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `jumlah_author_scholar` int(2) NOT NULL,
  `jumlah_paper_scholar` int(2) NOT NULL,
  `jumlah_sitasi_scholar` int(2) NOT NULL,
  `jumlah_jurnal_artikel` int(2) NOT NULL,
  `jumlah_book_chapter` int(2) NOT NULL,
  `jumlah_paper_conference` int(2) NOT NULL,
  `jumlah_artikel_scopus` int(2) NOT NULL,
  `jumlah_non_artikel_scopus` int(2) NOT NULL,
  `jumlah_sitasi_scopus` int(2) NOT NULL,
  `sinta_score` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=4006 DEFAULT
CHARSET=latin1;
```

Kode Sumber 4.2-9 Create Table jumlah\_sitasi\_sinta

#### 4.2.10 Create Tabel “jumlah\_author\_sinta”

```
CREATE TABLE `jumlah_author_sinta` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `id_data` int(2) NOT NULL,
  `nomor` int(2) NOT NULL,
  `nama_author` varchar(150) NOT NULL,
  `nama_universitas` varchar(150) NOT NULL,
  `jumlah_sitasi_scholar` int(2) NOT NULL,
  `i10_index_scholar` int(2) NOT NULL,
  `h_index_scholar` int(2) NOT NULL,
  `waktu_akses` tinytext NOT NULL,
```

### Kode Sumber 4.2-10 Create Table jumlah\_author\_sinta

#### 4.2.11 Create Tabel “precense”

```
CREATE TABLE `precense` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `url_web` varchar(255) NOT NULL,
  `domain_search` varchar(150) NOT NULL,
  `hasil_search` int(2) NOT NULL,
  `tanggal_akses` datetime NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=526 DEFAULT
CHARSET=latin1;
```

### Kode Sumber 4.2-11 Create Table precense

#### 4.2.12 Create Tabel “impact”

```
CREATE TABLE `impact` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `url_web` varchar(255) NOT NULL,
  `domain_web` varchar(150) NOT NULL,
  `page_authority` int(2) NOT NULL,
  `domain_authority` int(2) NOT NULL,
  `backlink_eksternal` int(2) NOT NULL,
  `tanggal_akses` datetime NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=1960 DEFAULT
```

### Kode Sumber 4.2-12 Create Table impact

#### 4.2.13 Create Tabel “alexa\_rank”

```
CREATE TABLE `alexa_rank` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `url_web` varchar(255) NOT NULL,
  `rank_world` int(2) NOT NULL,
  `rank_country` int(2) NOT NULL,
  `country_name` varchar(150) NOT NULL,
  `tanggal_akses` datetime NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=444 DEFAULT
CHARSET=latin1;
```

### Kode Sumber 4.2-13 Create Tabel alexa\_rank

#### 4.2.14 Create Tabel “kumpulan skor final”

```
CREATE TABLE `kumpulan_skor_final` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `kode_universitas` int(3) NOT NULL,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_reputasi_akademik_normalisasi` float NOT
NULL,
  `skor_nilai_akreditasi_normalisasi` float NOT
NULL,
  `skor_jumlah_mahasiswa_normalisasi` float NOT
NULL,
  `skor_jumlah_sitasi_normalisasi` float NOT NULL,
  `skor_jumlah_tenaga_pengajar_normalisasi` float
NOT NULL,
  `skor_kualitas_sitasi_normalisasi` float NOT
```

#### 4.2.15 Create Tabel “4icu”

```
CREATE TABLE `4icu` (  
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,  
  `nama_universitas` varchar(255) NOT NULL,  
  `skor_4icu` float NOT NULL,  
  `peringkat_4icu` int(2) NOT NULL,  
  PRIMARY KEY (`id`)  
  ) ENGINE=InnoDB AUTO_INCREMENT=2128 DEFAULT  
  CHARSET=latin1;
```



### Kode Sumber 4.2-16 Create Table 4icu

#### 4.2.16 Create Tabel “dikti”

```
CREATE TABLE `dikti` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_dikti` float NOT NULL,
  `peringkat_dikti` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=1773 DEFAULT
CHARSET=latin1;
```

### Kode Sumber 4.2-17 Create Table dikti

#### 4.2.17 Create Tabel “qs\_general”

```
CREATE TABLE `qs_general` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_qs` float NOT NULL,
  `peringkat_qs` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=2838 DEFAULT
CHARSET=latin1;
```

### Kode Sumber 4.2-18 Create Table qs\_general

#### 4.2.18 Create Tabel “qs\_asia”

```
CREATE TABLE `qs_asia` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_qs_asia` float NOT NULL,
  `peringkat_qs_asia` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=2838 DEFAULT
CHARSET=latin1;
```

### Kode Sumber 4.2-19 Create Table qs\_asia

#### 4.2.19 Create Tabel “qs\_plsa”

```
CREATE TABLE `qs_plsa` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_qs_plsa` float NOT NULL,
  `peringkat_qs_plsa` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=3193 DEFAULT
CHARSET=latin1;
```

Kode Sumber 4.2-20 Create Table qs\_plsa

#### 4.2.20 Create Tabel “webometrics”

```
CREATE TABLE `webometrics` (
  `id` int(2) unsigned NOT NULL AUTO_INCREMENT,
  `nama_universitas` varchar(255) NOT NULL,
  `skor_webometrics` float NOT NULL,
  `peringkat_webometrics` int(2) NOT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB AUTO_INCREMENT=2838 DEFAULT
CHARSET=latin1;
```

Kode Sumber 4.2-21 Create Table webometrics

### 4.3 Tahap Prapemrosesan

Pada bagian ini terdapat beberapa subproses di dalamnya. Semua proses yang terjadi merupakan proses pengolahan dokumen pada tahap awal. Secara garis besar terdapat beberapa macam

proses yang ada dalam prapemrosesan, yaitu pengambilan kata-kata, proses tokenisasi dan perhitungan frekuensi kata.

#### 4.3.1 Preprocessing

Sebelum memasukkan dokumen yang akan dilakukan proses sPLSA dilakukan tahap Preprocessing. Pada tahap ini merupakan tahap awal karena sangat berpengaruh pada hasil yang akan diuji coba. Tahap preprocessing yaitu tokenisasi dan *Stopword* (membuang kata yang tidak penting)

```

1. public static List<String> tokenizeSents(String sents){
2.     Reader reader = new StringReader(sents);
3.     DocumentPreprocessor dp = new DocumentPreprocessor(reader);
4.
5.     List<String> sentenceList = new LinkedList<String>();
6.     Iterator<List<HasWord>> it = dp.iterator();
7.     while (it.hasNext()) {
8.         StringBuilder sentenceSb = new StringBuilder();
9.         List<HasWord> sentence = it.next();
10.        for (HasWord token : sentence) {
11.            if(sentenceSb.length()>1) {
12.                sentenceSb.append(" ");
13.            }
14.            sentenceSb.append(token);
15.        }
16.        sentenceList.add(sentenceSb.toString());
17.    }
18.    return sentenceList;
19. }
20. for(int i = 0; i < words.size(); i++){
21.     if(Stopwords.isStopword(words.get(i)) || isNoiseWord(words.get(i))){
22.         words.remove(i);
23.         i--;
24.     }
25. }

```

### Kode Sumber 4.3-1 Kode Sumber Preprocessing

#### 4.3.2 Membaca Dokumen

Pada proses ini merupakan proses membaca dokumen dari 1 dokumen besar yang ada dalam 1 *folder* universitas.

```

1. public void readDocs(String docsPath){
2.     File[] files = new File(docsPath).listFiles();
3.     Arrays.sort(files, new AlphanumFileComparator());
4.     for(File docFile : files){
5.         Document doc = new Document(docFile.getAbsolutePath(), termToIndexMap, indexToTermMap, termCountMap);
6.         docs.add(doc);
7.     }
8. }

```

### Kode Sumber 4.3-2 Kode Sumber Membaca Dokumen

#### 4.3.3 Membaca Kata dalam Dokumen

Pada tahap ini membaca setiap kata yang ada pada dokumen besar, setiap kata yang dibaca adalah kata yang telah dilakukan *preprocessing* terlebih dahulu seperti membuang punctuation, tokenisasi (memisahkan kata) dan melakukan stopwords yaitu membuang kata yang tidak penting

```

1. public static class Document {
2.     private String docName;
3.     int[] docWords;
4.
5.     public Document(String docName, Map<String, Integer> termToIndexMap, ArrayList<String> indexToTermMap, Map<String, Integer> termCountMap){
6.         this.docName = docName;

```

```

7.          //Read file and initialize word index a
   rray
8.          ArrayList<String> docLines = new ArrayL
   ist<String>();
9.          ArrayList<String> words = new ArrayList
   <String>();
10.         List<String> tokenSents = new ArrayList
   <String>();
11.         FileUtil.readLines(docName, docLines);

12.         System.out.println(docName)
   ;
13.         for(String line : docLines){
14.             tokenSents = StanfordTokenizer.toke
   nizeSents(line);
15.             for(String tokenSent : tokenSents){
16.                 FileUtil.tokenizeAndLowerCase(t
   okenSent, words);
17.             }
18.         }

```

**Kode Sumber 4.3-3 Kode Sumber Membaca Kata dalam Dokumen**

#### 4.4 Tahap Perhitungan Probabilitas Topik terhadap Dokumen dan Kata

Pada tahap ini dilakukan perhitungan topik terhadap dokumen dan topik terhadap kata. Nilai probabilitas yang dihasilkan diurutkan dari jumlah topik yang ditentukan.

```

1. //init  $p(z|d)$ , for each document the constraint is s
   um( $p(z|d)$ )=1.0
2.     for (int i = 0; i < N; i++) {
3.         double[] pros = randomProbilities(topicN
   um);
4.         for (int j = 0; j < topicNum; j++) {
5.             docTopicPros[i][j] = pros[j];
6.         }
7.     }
8.     //init  $p(w|z)$ , for each topic the constraint
   is sum( $p(w|z)$ )=1.0

```

```

9.         for (int i = 0; i < topicNum; i++) {
10.            double[] pros = randomProbabilities(M);
11.            for (int j = 0; j < M; j++) {
12.                topicTermPros[i][j] = pros[j];
13.            }

```

**Kode Sumber 4.4 Kode Sumber Perhitungan Probabilitas Topik terhadap Dokumen dan Kata**

#### 4.5 Tahap Perhitungan Random Probability

Pada tahap ini metode PLSA mencari probabilitas terhadap dokumen dengan melakukan probabilitas secara acak

```

1. public double[] randomProbabilities(int size) {
2.     if (size < 1) {
3.         throw new IllegalArgumentException("The
   size param must be greate than zero");
4.     }
5.     double[] pros = new double[size];
6.
7.     int total = 0;
8.     Random r = new Random(2);
9.     for (int i = 0; i < pros.length; i++) {
10.        //avoid zero
11.        pros[i] = r.nextInt(size) + 1;
12.
13.        total += pros[i];
14.    }
15.
16.    //normalize
17.    for (int i = 0; i < pros.length; i++) {
18.        pros[i] = pros[i] / total;
19.    }
20.
21.    return pros;
22. }

```

**Kode Sumber 4.5 Kode Sumber Perhitungan Random Probability**

#### 4.6 Tahap Perhitungan *Expectation Maximization*

Pada tahap ini merupakan tahap terakhir dari metode PLSA yaitu menggunakan algoritma *Expectation Maximization* yaitu berfungsi untuk mendapatkan nilai yang *konvergen* dan menghentikan nilai random yang dilakukan pada tahap sebelumnya.

```

1.  /**
2.   *
3.   * Plsa inference using EM
4.   *
5.   * @param docs all documents
6.   * @throws IOException
7.   */
8.  public void inferenceModel(Documents docSet) throws IOException {
9.      //use em to estimate params
10.     //save model according to model parameters
11.
12.     if (iters < saveStep + beginSaveIters) {
13.         System.err.println("Error: the number of iterations
14.             should be larger than " + (saveStep + beginSaveIters));
15.         System.exit(0);
16.     }
17.     for (int i = 0; i < iters; i++) {
18.         System.out.println("Iteration " + i + " --
19.             -----");
20.         if ((i >= beginSaveIters)
21.             && ((i - beginSaveIters) % saveStep) == 0) {
22.             // Saving the model
23.             System.out.println("Saving model at iteration " + i + " ... ");
24.             saveIteratedModel(i, docSet);
25.         }
26.         em();

```

```

26.         System.out.println("After E step and M s
tep, the new log likelihood is " + computeLogLikeli
hood(docSet));
27.     }
28.
29.         System.out.println("done");
30.     }
31.
32.     private double computeLogLikelihood(Documents do
cSet) {
33.         // TODO Auto-generated method stub
34.         /*
35.          * Compute the log likelihood of generation
the corpus
36.          *
37.          *  $L = \sum_i \{n(d_i) * [\sum_j (n(d_i, w_j) /
n(d_i) * \log \sum_k p(z_k|d_i) * p(w_j|z_k))]\}$ 
38.          *
39.          */
40.         double L = 0.0;
41.         for (int i = 0; i < N; i++) {
42.             double docISize = docSet.docs.get(i).docWo
rds.length;
43.             double sumM = 0.0;
44.             for (int j = 0; j < M; j++) {
45.                 double sumK = 0.0;
46.                 for(int k = 0; k < topicNum; k++){
47.                     sumK += docTopicPros[i][k] * topic
TermPros[k][j];
48.                 }
49.                 //System.out.println("sumK: " + sumK);
50.                 sumM += (double)docTermMatrix[i][j] /
docISize * Math.log10(sumK);
51.             }
52.             //System.out.println("sumM: " + sumM);
53.             L += docISize * sumM;
54.         }
55.         return L;
56.     }
57.
58. /**
59.     *

```





```

91.         }
92.     }
93.
94.     //M-step
95.     /*
96.      * update  $p(w|z), p(w|z) = \text{sum}(n(d', w) * p(z|d', w$ 
97.      *  $, \&)) / \text{sum}(\text{sum}(n(d', w') * p(z|d', w', \&)))$ 
98.      *  $d'$  represent all documents
99.      *  $w'$  represent all vocabularies
100.     */
101.     /*
102.     */
103.     for (int topicIndex = 0; topicIndex <
        topicNum; topicIndex++) {
104.         double totalDenominator = 0.0;
105.         for (int wordIndex = 0; wordIndex
            < M; wordIndex++) {
106.             double numerator = 0.0;
107.             for (int docIndex = 0; docInd
                ex < N; docIndex++) {
108.                 numerator += docTermMatri
                    x[docIndex][wordIndex]
109.                     * docTermTopicPro
                        s[docIndex][wordIndex][topicIndex];
110.             }
111.             topicTermPros[topicIndex][wor
                dIndex] = numerator;
112.
113.             totalDenominator += numerator
114.         }
115.     }
116.
117.     if (totalDenominator == 0.0) {
118.         totalDenominator = avoidZero(
            totalDenominator);
119.     }
120.
121.     for (int wordIndex = 0; wordIndex
        < M; wordIndex++) {
122.         topicTermPros[topicIndex][wor
            dIndex] = topicTermPros[topicIndex][wordIndex]

```

```

123.                                     / totalDenominator;
124.                                     }
125.                                 }
126.                                /*
127.                                * update  $p(z|d), p(z|d) = \frac{\sum(n(d, w') * p(z|d, w'))}{\sum(\sum(n(d, w') * p(z'|d, w', &)))}$ 
128.                                *
129.                                *  $w'$  represent all vocabularies
130.                                *  $z'$  represnet all topics
131.                                *
132.                                */
133.                                for (int docIndex = 0; docIndex < N;
docIndex++) {
134.                                    //actually equal sum(w) of this d
oc
135.                                    double totalDenominator = 0.0;
136.                                    for (int topicIndex = 0; topicInd
ex < topicNum; topicIndex++) {
137.                                        double numerator = 0.0;
138.                                        for (int wordIndex = 0; wordI
ndex < M; wordIndex++) {
139.                                            numerator += docTermMatri
x[docIndex][wordIndex]
140.                                                * docTermTopicPro
s[docIndex][wordIndex][topicIndex];
141.                                        }
142.                                        docTopicPros[docIndex][topicI
ndex] = numerator;
143.                                        totalDenominator += numerator
;
144.                                    }
145.
146.                                    if (totalDenominator == 0.0) {
147.                                        totalDenominator = avoidZero(
totalDenominator);
148.                                    }
149.
150.                                    for (int topicIndex = 0; topicInd
ex < topicNum; topicIndex++) {
151.                                        docTopicPros[docIndex][topicI
ndex] = docTopicPros[docIndex][topicIndex]
152.                                            / totalDenominator;
153.                                    }

```

```

154.         }
155.     }

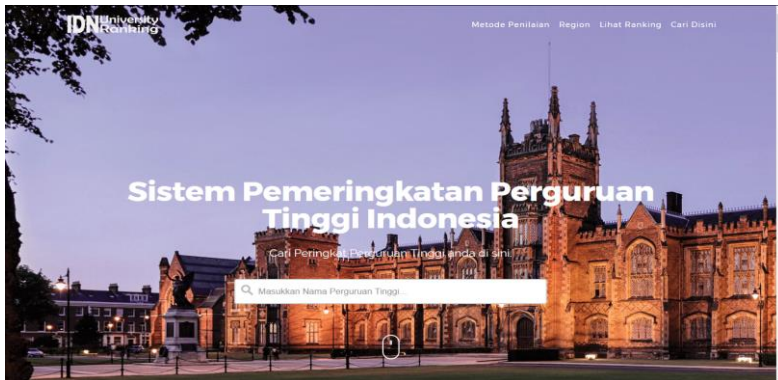
```

#### Kode Sumber 4.6 Kode Sumber Expectation Maximization

## 4.7 Implementasi Antarmuka Pengguna

### 4.7.1 Antarmuka Home (index)

Antarmuka ini sebagai halaman root dari sistem pemeringkatan yang muncul pertama kali ketika pengguna membuka domain sistem pada web browser.



**Gambar 4.7-1 Antarmuka Home**

### 4.7.2 Antarmuka Mencari Perguruan Tinggi Berdasarkan Potongan Nama

Pada antarmuka ini pengguna dapat mencari perguruan tinggi berdasarkan keyword potongan nama perguruan tinggi yang

diinputkan. Halaman ini juga menjadi halaman home dalam sistem perangkat lunak.



Gambar 4.7-2 Antarmuka Pencarian Nama

4.7.3 Antarmuka Hasil Pencarian Perguruan Tinggi Berdasarkan Keyword Potongan Nama

Pada antarmuka ini pengguna End User dapat melihat hasil pencarian perguruan tinggi berdasarkan keyword potongan nama.

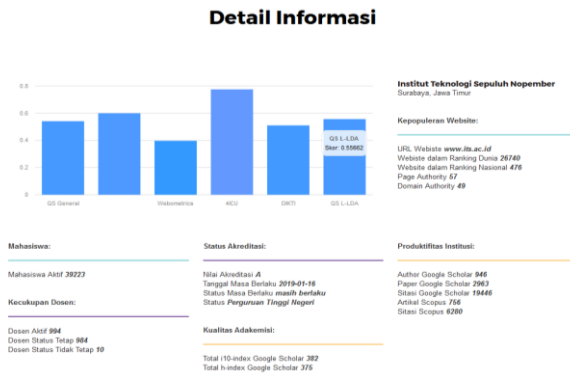
Hasil Pencarian

No.	Nama Perguruan Tinggi	Kota	Provinsi	QS General	QS Asia	Webometrics	4ICU	Dikti	QS L-LDA	Lainnya
1	Institut Teknologi Sepuluh Nopember	Surabaya	Jawa Timur	Peringkat 6	Peringkat 7	Peringkat 6	Peringkat 12	Peringkat 3	Peringkat 11	Lihat Detail

Gambar 4.7-3 Antarmuka Hasil Pencarian

4.7.4 Antarmuka Melihat Detail Informasi Perguruan Tinggi

Pada antarmuka ini pengguna dapat melihat informasi secara detail, diantaranya grafik peringkat perguruan tinggi berdasar tipe pemeringkatan, informasi mahasiswa, kecukupan dosen, status akreditasi, produktifitas akademisi, produktifitas institusi dan kepopuleran website.



**Gambar 4.7-4 Antarmuka Detail Informasi**

#### 4.7.5 Antarmuka Melihat Latar Belakang Sistem Pemeringkatan

Pada antarmuka ini pengguna dapat melihat latar belakang dari sistem pemeringkatan.



**Gambar 4.7-5 Antarmuka Informasi Latar Belakang**

4.7.6 Antarmuka Melihat Informasi Metode Penilaian

serta informasi mengenai kriteria metode yang digunakan dalam sistem pemerinkkatan ini.



Gambar 4.7-6 Antarmuka Informasi Metode Penilaian



Gambar 4.7-7 Antarmuka Metode Penilaian QS General



**Gambar 4.7-8 Antarmuka Metode Penilaian QS Asia**



**Gambar 4.7-9 Antarmuka Metode Penilaian Webometrics**

#### 4.7.7 Antarmuka Pencarian Peringkat Perguruan Tinggi Berdasarkan Region

Pada antarmuka ini pengguna dapat melihat peringkat perguruan tinggi berdasarkan region wilayah provinsi. Terdapat map wilayah Indonesia sebagai masukan pemilihan region wilayah yang ditentukan.





Gambar 4.7-10 Antarmuka Pencarian Berdasar Region

4.7.8    **Antarmuka Hasil Pencarian Perguruan Tinggi Berdasar Region Wilayah Perguruan Tinggi**

Pada antarmuka ini pengguna dapat melihat hasil peringkat dalam region wilayah perguruan tinggi.

**Peringkat Perguruan Tinggi**  
(Region: Jawa Timur)

QS General   QS Asia   Webometrics   4ICU   **Pemeringkatan Dikti**   QS L-LDA

Masukkan Nama Kota Diini...

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Universitas Brwajaya	0.561164	Malang	Jawa Timur
2	Universitas Airlangga	0.559164	Surabaya	Jawa Timur
3	Institut Teknologi Sepuluh Nopember	0.541009	Surabaya	Jawa Timur
4	Universitas Muhammadiyah Malang	0.527906	Malang	Jawa Timur
5	Universitas Kristen Petra	0.488639	Surabaya	Jawa Timur
6	Universitas Surabaya	0.472489	Surabaya	Jawa Timur
7	Universitas Muhammadiyah Sidoarjo	0.466495	Sidoarjo	Jawa Timur
8	Universitas Widya Kartika	0.462224	Surabaya	Jawa Timur
9	Universitas Merdeka Malang	0.456248	Malang	Jawa Timur
10	Universitas Trunojoyo	0.449357	Bangkalan	Jawa Timur

Gambar 4.7-11 Antarmuka Hasil Pencarian Berdasar Region

4.7.9    **Antarmuka Pembobotan Pemeringkatan Perguruan Tinggi**

Pada antarmuka ini pengguna dapat menentukan bobot setiap kriteria perguruan tinggi sesuai dengan kebutuhan. Terdapat

beberapa tipe pemeringkatan yang dapat digunakan untuk menghasilkan peringkat.

**IDN** University Ranking
Metode Penilaian   Region   Lihat Ranking   Cari Disini

### Atur Bobot Pemeringkatan

Pembobotan Pemeringkatan									
Indikator	QS General	QS Asia	Webometrics	4ICU	DIKTI	QS (L-LDA)	QS (PLSA)	QS (Helmholtz)	QS (Adaboost)
Reputasi Akademik	<input type="text" value="40"/>	<input type="text" value="30"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="25"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Reputasi Employer	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="14"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Mahasiswa Aktif	<input type="text" value="20"/>	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Sitasi Akademik	<input type="text" value="20"/>	<input type="text" value="10"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Tenaga Pengajar	<input type="text" value="5"/>	<input type="text" value="2.5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kualitas Sitasi (H-index & I10-index)	<input type="text" value="5"/>	<input type="text" value="2.5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="3"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Phd (S3)	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="20"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Paper Akademik	<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="2"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kecukupan Dosen	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="18"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kualitas Manajemen	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="30"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kualitas Kegiatan Kemahasiswaan	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kualitas Kegiatan Penelitian	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="30"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Total	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
	<b>Cek QS-General</b>	<b>Cek QS-Asia</b>	<b>Cek Webometrics</b>	<b>Cek 4ICU</b>	<b>Cek DIKTI</b>	<b>Cek QS L-LDA</b>	<b>Cek QS PLSA</b>	<b>Cek QS Helmholtz</b>	<b>Cek QS Adaboost</b>

**Gambar 4.7-12 Antarmuka Pembobotan Pemeringkatan**

#### 4.7.10 Antarmuka Hasil Pemeringkatan Perguruan Tinggi dengan Berbagai Macam Tipe Pemeringkatan

Pada antarmuka ini pengguna dapat melihat hasil pemeringkatan perguruan tinggi dari bobot yang sudah ditentukan pada proses pembobotan. Terdapat hasil pemeringkatan dari berbagai macam tipe pemeringkatan, diantaranya QS General, QS Asia, Webometrics, 4ICU dan pemeringkatan DIKTI.

Peringkat Perguruan Tinggi  
(QS General)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi	Compare (gilih 3)
1	Institut Teknologi Bandung	0.6801	Bandung	Jawa Barat	<input type="checkbox"/>
2	Universitas Gadjah Mada	0.6522	Yogyakarta	DI Yogyakarta	<input type="checkbox"/>
3	Universitas Indonesia	0.6414	Depok	Jawa Barat	<input type="checkbox"/>
4	Institut Pertanian Bogor	0.6673	Bogor	Jawa Barat	<input type="checkbox"/>
5	Universitas Diponegoro	0.5623	Semarang	Jawa Tengah	<input type="checkbox"/>
6	Universitas Brawijaya	0.5612	Malang	Jawa Timur	<input type="checkbox"/>
7	Universitas Airlangga	0.5592	Surabaya	Jawa Timur	<input type="checkbox"/>
8	Institut Teknologi Sepuluh Nopember	0.5411	Surabaya	Jawa Timur	<input type="checkbox"/>
9	Universitas Muhammadiyah Malang	0.5279	Malang	Jawa Timur	<input type="checkbox"/>
10	Universitas Padjadjaran	0.5224	Bandung	Jawa Barat	<input type="checkbox"/>

Gambar 4.7-13 Antarmuka Hasil Pemeringkatan QS General

Peringkat Perguruan Tinggi  
(QS Asia)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Institut Teknologi Bandung	0.7626	Bandung	Jawa Barat
2	Universitas Indonesia	0.7286	Depok	Jawa Barat
3	Universitas Gadjah Mada	0.7235	Yogyakarta	DI Yogyakarta
4	Institut Pertanian Bogor	0.6598	Bogor	Jawa Barat
5	Universitas Brawijaya	0.6165	Malang	Jawa Timur
6	Universitas Diponegoro	0.6053	Semarang	Jawa Tengah
7	Institut Teknologi Sepuluh Nopember	0.5995	Surabaya	Jawa Timur
8	Universitas Airlangga	0.5992	Surabaya	Jawa Timur
9	Universitas Hasanuddin	0.5671	Makassar	Sulawesi Selatan
10	Universitas Padjadjaran	0.5666	Bandung	Jawa Barat

Gambar 4.7-14 Antarmuka Hasil Pemeringkatan QS Asia

Peringkat Perguruan Tinggi  
(Webometrics)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Universitas Gadjah Mada	0.8207	Yogyakarta	DI Yogyakarta
2	Institut Teknologi Bandung	0.6314	Bandung	Jawa Barat
3	Universitas Diponegoro	0.5234	Semarang	Jawa Tengah
4	Universitas Sumatera Utara	0.4745	Medan	Sumatera Utara
5	Institut Pertanian Bogor	0.4659	Bogor	Jawa Barat
6	Universitas Indonesia	0.4468	Depok	Jawa Barat
7	Universitas Negeri Yogyakarta	0.4034	Yogyakarta	DI Yogyakarta
8	Institut Teknologi Sepuluh Nopember	0.3968	Surabaya	Jawa Timur
9	Universitas Brawijaya	0.3843	Malang	Jawa Timur
10	Universitas Padjadjaran	0.3605	Bandung	Jawa Barat

Gambar 4.7-15 Antarmuka Hasil Pemeringkatan Webometrics

Peringkat Perguruan Tinggi  
(4ICU)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Universitas Sumatera Utara	0.9213	Medan	Sumatera Utara
2	Institut Teknologi Bandung	0.834	Bandung	Jawa Barat
3	Universitas Indonesia	0.8263	Depok	Jawa Barat
4	Universitas Gadjah Mada	0.8205	Yogyakarta	DI Yogyakarta
5	Institut Pertanian Bogor	0.8105	Bogor	Jawa Barat
6	Universitas Airlangga	0.7911	Surabaya	Jawa Timur
7	Universitas Padjadjaran	0.7858	Bandung	Jawa Barat
8	Universitas Diponegoro	0.7839	Semarang	Jawa Tengah
9	Universitas Pendidikan Indonesia	0.7788	Bandung	Jawa Barat

Gambar 4.7-16 Antarmuka Hasil Pemeringkatan 4ICU

Peringkat Perguruan Tinggi  
(Pemeringkatan DIKTI)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Institut Teknologi Bandung	0.6988	Bandung	Jawa Barat
2	Institut Pertanian Bogor	0.5355	Bogor	Jawa Barat
3	Institut Teknologi Sepuluh Nopember	0.6161	Surabaya	Jawa Timur
4	Universitas Indonesia	0.4727	Depok	Jawa Barat
5	Universitas Gadjah Mada	0.4449	Yogyakarta	DI Yogyakarta
6	Universitas Negeri Jakarta	0.4188	Jakarta	DKI Jakarta
7	Universitas Diponegoro	0.406	Semarang	Jawa Tengah
8	Universitas Hasanuddin	0.4009	Makassar	Sulawesi Selatan

Gambar 4.7-17 Antarmuka Hasil Pemeringkatan DIKTI

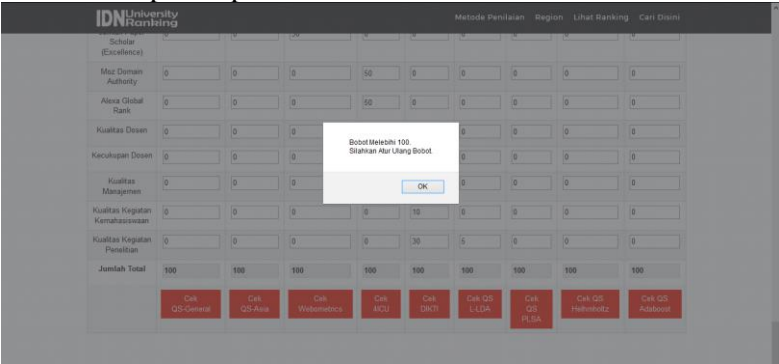
Peringkat Perguruan Tinggi  
(QS pLSA)

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1	Institut Teknologi Bandung	0.7434	Bandung	Jawa Barat
2	Universitas Indonesia	0.7123	Depok	Jawa Barat
3	Universitas Gadjah Mada	0.68	Yogyakarta	DI Yogyakarta
4	Institut Pertanian Bogor	0.6154	Bogor	Jawa Barat
5	Universitas Diponegoro	0.5877	Semarang	Jawa Tengah
6	Institut Teknologi Sepuluh Nopember	0.5841	Surabaya	Jawa Timur
7	Universitas Brawijaya	0.5764	Malang	Jawa Timur
8	Universitas Padjadjaran	0.5755	Bandung	Jawa Barat
9	Universitas Airlangga	0.5617	Surabaya	Jawa Timur
10	Universitas Sebelas Maret	0.5373	Surakarta	Jawa Tengah
11	Universitas Muhammadiyah Surakarta	0.5329	Surakarta	Jawa Tengah
12	Universitas Bina Nusantara	0.5243	Jakarta	DKI Jakarta

Gambar 4.7-18 Antarmuka Hasil Peringkat QS PLSA

4.7.11 **Antarmuka Error Handling Pembobotan Kriteria Pemingkatan**

Pada antarmuka ini pengguna dapat melihat pemberitahuan bahwa bobot yang diharuskan untuk diinputkan sebagai masukan pembobotan adalah 100, jika melebihi 100 maka akan menampilkan pemberitahuan.



**Gambar 4.7-19 Antarmuka Error Handling Pembobotan**

## **BAB V**

### **PENGUJIAN DAN EVALUASI**

Pada bab ini dibahas mengenai hasil uji coba perangkat lunak yang telah dibuat. Uji coba dilakukan untuk mengetahui kinerja sistem dengan lingkungan uji coba yang telah ditentukan dan menguji apakah fungsionalitas aplikasi telah diimplementasikan dengan benar dan berjalan sebagaimana mestinya.

#### **5.1 Lingkungan Uji Coba**

Lingkungan uji coba ini menjelaskan lingkungan yang digunakan untuk menguji implementasi dari pembuatan perangkat lunak pada Tugas Akhir. Lingkungan uji coba meliputi perangkat keras dan perangkat lunak yang dijelaskan sebagai berikut:

1. Perangkat keras
  - a. Prosesor: Intel® Core™ i7-5500U CPU @2.40GHz (4CPUs)
  - b. Memori (RAM): 12 GB
  - c. Hardisk (HDD): 1 TB
  - d. Tipe sistem: 64-bit sistem operasi
2. Perangkat lunak
  - a. Sistem operasi: Windows 8.1 Profesional x64 (64-bit)
  - b. Perangkat pengembang: Python v2.7.12, Anaconda 4.2.0 (64-bit), dan SublimeText 2 v2.0.

#### **5.2 Data Uji Coba**

Data uji coba digunakan untuk menguji jalannya sistem apakah sudah sesuai dengan tujuan dan lingkup pengerjaan yang ditujukan. Data uji coba yang digunakan dalam pengembangan perangkat lunak terdiri dari beberapa macam, diantaranya:

##### **5.2.1 Data Dokumen Abstrak Paper Akademisi**

Data uji coba yang digunakan dalam proses teks mining menggunakan metode PLSA adalah dokumen abstrak *paper*

akademisi dari 16 perguruan tinggi yang sesuai dengan Ground Truth reputasi akademik *QS World University rankings* 2015-2016. Penggunaan 16 perguruan tinggi ini didasarkan pada kepopuleran dan kualitas akademisi perguruan tinggi sesuai dengan standart *QS World University rankings* [4]. Tidak semua perguruan tinggi dapat masuk dalam peringkat *QS World University rankings*, baik secara general maupun secara terpisah dari kriteria-kriteria penilaiannya.

Daftar 16 perguruan tinggi berdasarkan peringkat kriteria reputasi akademik *QS World University Rankings* 2015-2016 seperti pada Tabel 5.2-1.

**Tabel 5.2-1 Peringkat Perguruan Tinggi**

No.	Perguruan Tinggi	Peringkat
1.	Institut Teknologi Bandung	1
2.	Universitas Indonesia	2
3.	Universitas Gadjah Mada	3
4.	Universitas Muhammadiyah Semarang	4
5.	Universitas Airlangga	5
6.	Institut Pertanian Bogor	6
7.	Univeristas Diponegoro	7
8.	Universitas Brawijaya	8
9.	Institut Teknologi Sepuluh Nopember	9
10.	Universitas Padjadjaran	10
11.	Universitas Islam Indonesia	11
12.	Univeritas 17 Agustus 1945	12
13.	Universitas Bina Nusantara	13
14.	Universitas Pelita Harapan	14
15.	Universitas Sebelas Maret	15
16.	Universitas Mataram	16

Dari 16 perguruan tinggi diatas, digunakan 50 dokumen abstrak *paper* dari masing-masing perguruan tinggi dan dihasilkan 800 buah dokumen dengan total kata yang dimiliki sejumlah 139919 buah kata. Data ini yang selanjutnya akan disebut dengan dataset abstrak *paper*.

Kata-kata ini akan diproses lebih lanjut dengan tahap preprocessing, untuk menganalisa duplikasi kata, karakter non alfabet, dan banyak frekuensi kata yang tidak berkepentingan yang ada di dalam dokumen. Tahap preprocessing yang digunakan meliputi case folding, tokenisasi, remove punctuation, dan stopwords removal [22]. Hasil reduksi jumlah kata dari semua dataset abstrak *paper* akademisi seperti pada Tabel 5.2-2.

**Tabel 5.2-2 Daftar Reduksi Jumlah Kata Dataset**

No.	Perguruan Tinggi	Kata Asli	Kata Hasil Preprocess
1.	Institut Pertanian Bogor	10539	6293
2.	Institut Teknologi Bandung	9287	5675
3.	Institut Teknologi Sepuluh Nopember	9630	6015
4.	Universitas 17 Agustus 1945	9146	5580
5.	Univeritas Airlangga	9457	6009
6.	Universitas Bina Nusantara	7236	4432
7.	Universitas Brawijaya	8651	5469
8.	Universitas Diponegoro	8397	5327
9.	Universitas Gadjah Mada	10214	6279
10.	Universitas Indonesia	8823	5645
11.	Universitas Islam Indonesia	7897	4989
12.	Universitas Mataram	7527	4561
13.	Universitas Muhammadiyah Surakarta	10330	6117
14.	Universitas Padjadjaran	7774	5008
15.	Universitas Pelita Harapan	7204	4542
16.	Universitas Sebelas Maret	7808	5244

Dari Tabel 5.2.-2, hasil reduksi pada tahap preprocessing didapatkan rata-rata 62.42% kata bersih dari total semua jumlah kata. Kata-kata bersih ini yang kemudian akan menjadi masukan untuk tahap selanjutnya.



### 5.2.2 Data Corpus Taxonomy Bloom

Selain penggunaan dataset abstrak paper, dalam menghasilkan proses teks mining topic modeling metode PLSA membutuhkan corpus data yang berisikan kumpulan kata kunci (*keyword*). Terdapat corpus Taxonomy Bloom yang sudah ada sebelumnya. Masing-masing corpus Taxonomy Bloom mempunyai jumlah keyword pada masing-masing level (6 level). Dalam Tugas Akhir ini mengacu pada beberapa referensi Taxonomy Bloom, diantaranya:

1. [www.cte.cornell.edu](http://www.cte.cornell.edu)
2. [www.bloomtaxonomy.org](http://www.bloomtaxonomy.org)
3. [www.teaching.uncc.edu](http://www.teaching.uncc.edu)
4. [www.web.stanford.edu](http://www.web.stanford.edu)

Dari masing-masing referensi Taxonomy Bloom diatas, didapatkan jumlah *keyword* seperti pada Tabel 5.2-3.

**Tabel 5.2-3 Rincian Jumlah Kata Corpus Taxonomy Bloom**

No.	Cte.cor nell.edu	Bloomtaxo nomy.org	(teaching.uncc.edu)			Web.stan ford.edu
			KC Metro	UMUC	GA Tech	
1.	31	21	21	9	10	8
2.	43	15	15	9	8	12
3.	47	16	15	9	9	12
4.	40	25	24	9	7	19
5.	52	34	36	7	10	14
6.	36	34	36	6	15	13

Keterangan:

1. Nomor 1, menyatakan keyword pada level Knowledge.
2. Nomor 2, menyatakan keyword pada level Comprehension.
3. Nomor 3, menyatakan kewyord pada level Application.
4. Nomor 4, menyatakan keyword pada level Analysis.
5. Nomor 5, menyatakan keyword pada level Synthesis.
6. Nomor 6, menyatakan keyword pada level Evaluation.

Dalam Tugas Akhir ini menggabungkan semua corpus Taxonomy Bloom dari ke-4 referensi yang sudah disebutkan. Memungkinkan ada keyword yang sama, dan kemudian akan direduksi dengan cara menghapus keyword yang sama. Tujuan dari menggabungkan adalah untuk memperkaya corpus Taxonomy Bloom sebagai dasar untuk membuat corpus Tingkat Kesiapterapan Teknologi (TKT).

Hasil penggabungan 4 referensi corpus Taxonomy Bloom dengan mereduksi keyword yang sama menjadi corpus taxonomy Bloom final seperti pada Tabel 5.2.-4.

**Tabel 5.2-4 Penggabungan Referensi Corpus Taxonomy Bloom**

No.	Level	Jumlah Kata
1.	Knowledge	35
2.	Comprehension	29
3.	Application	36
4.	Analysis	51
5.	Synthesis	51
6.	Evaluation	46

### 5.2.3 Data Corpus Tingkat Kesiapterapan Teknologi (TKT)

Data corpus TKT didapatkan dari beberapa cara yaitu dengan breakdown 6 level corpus Taxonomy Bloom, dan pengembangan menggunakan beberapa cara seperti teknik POS Tagging dan sinonim word. Tujuannya adalah untuk memperkaya *keyword* pada corpus TKT.

#### 5.2.3.1 Breakdown 6 Level Corpus Taxonomy Bloom

Dalam corpus Taxonomy Bloom terdapat keyword yang terdiri dari lebih dari satu kata, sehingga tidak dapat digunakan untuk pencocokan pada proses selanjutnya, maka dari itu keyword yang terdiri dari lebih dari satu kata akan dipecah menjadi satu-satu. Sehingga jumlah kata pada 6 level Corpus Taxonomy Bloom juga mengalami proses reduksi dan penambahan.

Dari hasil corpus Taxonomy Blom yang sudah mengandung kata yang berbeda di masing-masing level akan diurutkan berdasarkan tingkat makna konten kata. Kata yang mempunyai makna dasar berada pada urutan paling bawah, dan yang mempunyai makna implementasi/pengembangan berada pada urutan paling atas. Urutan tingkat makna kata ini akan dipecah menjadi 9 bagian sebagai implementasi terhadap corpus TKT yang mempunyai 9 level. Hasil breakdown menjadi 9 level corpus TKT seperti pada Tabel 5.2-5.

**Tabel 5.2-5 Hasil Breakdown Menjadi 9 Corpus TKT**

No.	Level TKT	Jumlah Kata
1.	TKT 1	31
2.	TKT 2	24
3.	TKT 3	32
4.	TKT 4	15
5.	TKT 5	31
6.	TKT 6	24
7.	TKT 7	26
8.	TKT 8	30
9.	TKT 9	34

#### 5.2.3.2 Pengembangan Menggunakan POS Tagging

Hasil corpus TKT menggunakan cara breakdown dari corpus Taxonomy Bloom memperoleh kata-kata dasar, dan jumlahnya masih cenderung sedikit bila dibandingkan dengan kata (*term*) pada dataset yang mencapai 139919 buah kata. Hal ini menjadi salah satu kunci hasil akurasi kecocokan kata. Dengan penambahan menggunakan POS Tagging diharapkan mampu menambah pengetahuan keyword dalam corpus TKT. Dilakukan beberapa percobaan untuk memperkaya corpus TKT dengan POS Tagging, diantaranya:

1. Akuisisi kata 1 (-1, +1): akuisisi kata naik satu kata dan turun satu kata.

2. Akuisisi kata 2 (-2, +2): akuisisi kata naik dua kata dan turun dua kata.
3. Akuisisi kata 3 (-3, +3): akuisisi kata naik tiga kata dan turun tiga kata.
4. Akuisisi kata 4 (-4, +4): akuisisi kata naik empat kata dan turun empat kata.
5. Akuisisi kata 8 (-5, +5): akuisisi kata naik delapan kata dan turun delapan kata.
6. Akuisisi kata 10 (-10, +10): akuisisi kata naik sepuluh kata dan turun sepuluh kata.

Hasil penambahan kata corpus TKT menggunakan POS Tagging dengan percobaan akuisisi seperti pada Tabel 5.2-6.

**Tabel 5.2-6 Hasil Penambahan Kata Corpus TKT Menggunakan POS Tagging**

Level	Jumlah Kata (default)	Jumlah Kata Akuisisi untuk TKT ke-					
		1	2	3	4	8	10
TKT 1	31	379	610	789	917	1313	1481
TKT 2	24	233	328	420	506	735	820
TKT 3	32	321	489	611	741	1032	1123
TKT 4	15	275	436	561	778	1010	1139
TKT 5	31	212	332	425	504	741	796
TKT 6	24	228	324	407	463	587	618
TKT 7	26	156	224	295	347	527	584
TKT 8	30	90	122	159	193	288	323
TKT 9	34	169	259	300	345	517	581

Keyword dalam masing-masing level corpus TKT hasil dari breakdown corpus Taxonomy Boolm dan teknik POS Tagging ini akan menjadi corpus TKT final dalam teks mining. Kecerdasan dokumen semakin bagus jika corpus yang dimiliki semakin banyak keywordnya.

### 5.2.3.3 Pengembangan Menggunakan Sinonim Word WordNet

Salah satu cara untuk memperkaya corpus adalah dengan mencari sinonim kata dari setiap kata yang ada pada 9 level corpus TKT. Cara ini digunakan sebagai cara pengganti dari penggunaan teknik POS Tagging, karena berdasar hasil analisa akhir menghasilkan hasil yang kurang memuaskan. Penggunaan sinonim word ini cukup efektif karena lebih menghemat waktu kerja dan juga lebih banyak kata yang bisa diserap dalam proses pencocokan kata daripada menggunakan teknik POS Tagging. Kata dari hasil penambahan sinonim ini akan ditambahkan pada level corpus TKT kata dasarnya. Sehingga menjadikan setiap level TKT menjadi kaya akan kata.

Dilakukan beberapa percobaan untuk memperkaya corpus menggunakan sinonim word, diantaranya:

1. Sinonim level 1: mencari satu sinonim teratas dari setiap kata dari ke-9 level TKT yang dilihat dari probabilitas kesamaan paling tinggi.
2. Sinonim level 2: mencari dua sinonim teratas yang dilihat dari dua probabilitas kesamaan paling tinggi.
3. Sinonim level 3: mencari tiga sinonim teratas yang dilihat dari tiga probabilitas kesamaan paling tinggi.
4. Sinonim level all: mengambil semua sinonim dari setiap kata dari ke-9 level TKT.

Hasil penambahan kata corpus TKT menggunakan sinonim word seperti pada Tabel 5.2-7.

**Tabel 5.2-7 Hasil Penambahan Kata Corpus TKT Menggunakan Sinonim Word**

Level	Jumlah Kata (default)	Total kata corpus TKT dengan Level sinonim word			
		1	2	3	all
TKT 1	31	50	57	62	132
TKT 2	24	43	54	59	75
TKT 3	32	41	44	47	70
TKT 4	15	25	29	30	50

TKT 5	31	55	68	73	120
TKT 6	24	28	29	29	31
TKT 7	26	48	58	63	122
TKT 8	30	55	66	72	109
TKT 9	34	60	74	77	100

Dari percobaan memperkaya corpus TKT diatas baik menggunakan teknik POS Tagging dan sinonim word, satu per satu diuji coba untuk melakukan pencocokan kata sebagai asumsi label awal dokumen. Disimpulkan bahwa hasil akhir yang paling bagus dari segi distance peringkat perguruan tinggi akhir adalah memperkaya corpus menggunakan sinonim word level all.

#### 5.2.4 Rekap Jumlah Mahasiswa dan Tenaga Pengajar

Data jumlah mahasiswa dan tenaga pengajar merupakan salah satu data utama untuk kriteria pemeringkatan. Data ini didapatkan dari database Pangkalan Data Pendidikan Tinggi (PDDIKTI). Pengguna umum dapat mengakses halaman pada alamat <http://www.forlap.dikti.go.id>. Data yang disediakan dapat diunduh menjadi file excel.

##### 5.2.4.1 Data Jumlah Mahasiswa

Data jumlah mahasiswa yang digunakan adalah mengacu pada data perguruan tinggi yang terdaftar pada PDDIKTI. Perguruan tinggi yang terdaftar meliputi Institut, Universitas, Sekolah Tinggi, Sekolah Tinggi Ilmu Kesehatan, Akademi, dan Politeknik.

Dalam Tugas Akhir ini dibatasi perguruan tinggi dalam lingkup Universitas dan Institut saja karena ranah ilmunya cukup merata. Terdapat 2 kelas yaitu perguruan tinggi negeri dan perguruan tinggi swasta yang dibagi menjadi 14 sub wilayah kopertis. Informasi yang akan digunakan untuk masukan sistem adalah dari beberapa atribut pada rekapan data, yaitu masing-masing perguruan tinggi akan diambil jumlah mahasiswa pada 2 tahun

terakhir dalam periode rekap. Misal untuk rekap tahun 2017, maka data yang digunakan adalah jumlah mahasiswa semester 1 dan 2 tahun 2016, dan semester 1 dan 2 tahun 2017. Rincian jumlah perguruan tinggi dalam rekap seperti pada Tabel 5.2-8.

**Tabel 5.2-8 Daftar Jumlah Perguruan Tinggi Berdasarkan Jenis Wilayah Administrasi**

No.	Jenis Perguruan Tinggi	Jumlah Perguruan Tinggi
1.	PTN	122
2.	Kopertis wilayah 1	266
3.	Kopertis wilayah 2	217
4.	Kopertis wilayah 3	333
5.	Kopertis wilayah 4	478
6.	Kopertis wilayah 5	105
7.	Kopertis wilayah 6	249
8.	Kopertis wilayah 7	330
9.	Kopertis wilayah 8	165
10.	Kopertis wilayah 9	364
11.	Kopertis wilayah 10	248
12.	Kopertis wilayah 11	168
13.	Kopertis wilayah 12	44
14.	Kopertis wilayah 13	109
15.	Kopertis wilayah 14	60

#### 5.2.4.2 Data Jumlah Tenaga Pengajar

Data jumlah tenaga pengajar mempunyai beberapa kriteria khusus yang disediakan oleh PDDIKTI, diantaranya rekap tenaga pengajar berdasarkan jabatan akademik, berdasarkan status kepegawaian, dan berdasarkan pendidikan tertinggi. Daftar perguruan tinggi yang digunakan sama seperti Tabel 5.2.4.1 yang hanya dipilih yang berbadan Institut atau Universitas saja. Informasi yang akan digunakan untuk masukan sistem dari beberapa atribut pada rekapan data, diantaranya:

1. Berdasarkan status pendidikan tertinggi
  - a) Jenjang diploma (D1, D2, D3, D4)
  - b) Jenjang sarjana (S1, S2, S3)
  - c) Jenjang spesialis (SP1, SP2)
  - d) Jenjang profesi
  - e) Jenang non-formal dan informal
2. Berdasarkan jabatan
  - a) Jabatan akademik lektor
  - b) Jabatan akademik lektor kepala
  - c) Jabatan akademik profesor
3. Berdasarkan status kepegawaian
  - a) Dosen dengan status tetap
  - b) Dosen dengan status tidak tetap

#### 5.2.5 Data Online Science and Technology Index (SINTA)

Data online dari Science and Technology Index (SINTA) berupa data yang berhubungan dengan jumlah penelitian akademisi perguruan tinggi. SINTA dapat diakses pada alamat [www.sinta1.ristekdikti.go.id](http://www.sinta1.ristekdikti.go.id). Data yang disediakan berupa informasi yang ditampilkan dalam konten website. Oleh karena itu dalam kriteria ini proses pengambilan data menggunakan teknik *grabbing* dari halaman website.

Pemilihan data online dari SINTA bertujuan untuk mengumpulkan informasi tentang penelitian akademisi yang mempunyai standart tinggi. Data yang akan diolah harus masuk dalam skala database Scopus sebagai pusat data penelitian terbaik saat ini. Data yang dihimpun oleh SINTA sesuai dengan kriteria keberadaan di Scopus [18].

Secara tidak langsung menggunakan teknik *grabbing* dari website SINTA untuk mengoptimasi pengolahan data yang berbasis pada Scopus, karena Scopus tidak memberikan akses secara mudah untuk dapat menggunakan databasenya. Sehingga penggunaan pihak ketiga sebagai penyedia data menjadi alternatif dalam penentuan data kriteria ini.



Berikut beberapa konten yang dapat diambil dari website SINTA dan digunakan dalam kriteria penilaian:

1. Authors, merupakan jumlah penulis penelitian yang ter-index oleh Google Scholar dan Scopus.
2. Google Citation, merupakan jumlah sitasi dari semua author dalam perguruan tinggi yang ter-index oleh Google Scholar.
3. Google Documents, merupakan jumlah dokumen penelitian (*paper*) yang ter-index Google Scholar.
4. Google Scholar Indexed, merupakan jumlah h-index dan i10-index dari setiap akademisi yang ter-index oleh Google Scholar.
5. Scopus Indexed, merupakan jumlah h-index dan i10-index dari setiap akademisi yang ter-index oleh Scopus.
6. Journal Articles, merupakan jumlah paper penelitian yang masuk dalam jurnal.
7. Articles in Scopus, merupakan artikel penelitian dari akademisi yang masuk dalam database Scopus.
8. Non Articles in Scopus, merupakan non artikel yang ter-index oleh database Scopus.

Sebagai catatan bahwa data yang disediakan oleh SINTA kurang valid karena tidak adanya pembaharuan pada database SINTA. Data SINTA yang digunakan pada tugas akhir ini merupakan data pada periode 2016-2017. Jumlah data pada masing masing variable pada SINTA perlu dilakukan validasi lagi mengikuti periode data SINTA pada setiap tahunnya.

Alasan menggunakan data dari SINTA karena data yang disediakan bisa diambil secara gratis dan sesuai dengan variable indikator penilaian QS World University Ranking yaitu sebagai jumlah sitasi. Jika data SINTA diperbaharui setiap periodenya, maka hasil dari sistem akan berubah sesuai dengan kondisi data yang *real time* dengan bobot indikator yaitu 20%

## 5.2.6 Data Online Analisa Webhost

Data analisa webhost merupakan data yang didapatkan dari teknik *grabbing* pada informasi *webhost*. Informasi yang didapatkan berfokus pada traffict website perguruan tinggi.

Penggunaan analisa *webhost* ini merupakan salah satu kriteria penilaian pada tipe pemeringkatan lain yaitu 4ICU dan Webometrics, yang sebgain besar kriteria masih bergantung pada informasi jan jumlah traffict website. Dalam Tugas Akhir ini, kriteria yang bergantung pada traffict website akan digunakan sebagai kriteria tambahan yang dapat digunakan oleh pengguna perangkat lunak jika ingin mengembangkan masukan kriteria penilaian.

Pengambilan data online ini memanfaatkan beberapa API tambahan untuk memaksimalkan hasil dan memudahkan analisa. Memanfaatkan API sangat dianjurkan untuk mengurangi running time dan detail dalam penyediaan informasi. API yang digunakan antara lain Alexa Rank API, dan MOZ Analytic API.

### 5.2.6.1 Alexa Rank

Alexa Rank API berfokus pada pengalihan informasi traffict rank dari *webhost* (domain, subdomain). Informasi yang didapatkan menggunakan Alexa Rank API antara lain:

1. World Rank, merupakan peringkat webhost dalam peringkat dunia. Peringkat ini sangat mempengaruhi kepopuleran sebuah website. Semakin tinggi peringkat, maka akan semakin kecil nominal rank yang didapatkan.
2. Country Rank, merupakan peringkat webhost dalam peringkat negara. Peringkat ini diperoleh dari analisa lokasi webhost berasal.
3. Country Name, merupakan informasi lokasi negara dimana webhost berasal.

### 5.2.6.2 Data Jumlah Webhost dari Google

Data jumlah webhost yang digunakan adalah jumlah domain utama, subdomain, dan repository dari website perguruan tinggi yang ter-index oleh Google. Pengambilan data ini menggunakan teknik *grabbing* dengan memanfaatkan API dari bahasa pemrograman PHP yaitu simpleHTMLDOM. Dimana sistem akan mengambil konten website pencarian Google, kemudian mengambil konten *results page* yang didapatkan, hasil *results page* ini yang akan dijumlah dan disimpan sebagai jumlah webhost. Keseluruhan data online analisa webhost, data rekap jumlah mahasiswa dan tenaga pengajar akan disimpan dalam database yang sudah dijelaskan pada Bab 3.

## 5.3 Skenario Pengujian

Dalam uji coba yang dilakukan dalam tugas akhir ini memiliki beberapa tahapan yang dijelaskan pada subbab ini.

### 5.3.1 Skenario Pengambilan Data Online Menggunakan Teknik *Grabbing*

Skenario pengambilan data online menggunakan jaringan dan koneksi internet untuk tersambung pada server maupun website yang akan diambil datanya.

#### 5.3.1.1 Skor Nilai Akreditasi

Metode	: Teknik <i>Grabbing</i>
Sasaran	: Penyedia informasi akreditasi perguruan tinggi <ul style="list-style-type: none"> <li>- <a href="https://banpt.or.id/direktori/institusi/pencarian_institusi">https://banpt.or.id/direktori/institusi/pencarian_institusi</a></li> <li>- <a href="http://daftarkampus.com/daftar-akreditasi-perguruan-tinggi-negeri-dan-swasta-lengkap/">http://daftarkampus.com/daftar-akreditasi-perguruan-tinggi-negeri-dan-swasta-lengkap/</a></li> </ul>
Masukan	: Halaman website pada poin sasaran

Keluaran : Hasil informasi dalam beberapa atribut

- nama\_universitas
- peringkat\_huruf
- masa\_berlaku
- status\_berlaku
- skor\_masa\_berlaku
- tmp\_nama\_universitas

Jumlah Data : 921 data

Tabel Database : tampung\_akreditasi

Tipe Tabel : Single table

Tipe Pemrograman : Bahasa PHP dan Library

Status Kebutuhan : Terpenuhi

#### 5.3.1.2 Skor Jumlah Sitasi

Metode : Teknik *Grabbing*

Sasaran : Penyedia informasi dari database SINTA Ristek

Dikti

- <http://sinta1.ristekdikti.go.id/>

Masukan : Halaman website pada poin sasaran

Keluaran : Hasil informasi dalam beberapa atribut

- nama\_universitas
- jumlah\_author\_scholar
- jumlah\_paper\_scholar
- jumlah\_sitasi\_scholar
- jumlah\_jurnal\_artikel
- jumlah\_book\_chapter
- jumlah\_paper\_conference
- jumlah\_artikel\_scopus
- jumlah\_non\_artikel\_scopus
- jumlah\_sitasi\_scopus
- sinta\_score

Jumlah Data : 3471 data

Tabel Database :

- tampung\_sitasi\_sinta

	-	tampung_sitasi_sinta_2
	-	tapung_log_sitasi_sinta
	-	tampung_sitasi_sinta_3
	-	jumlah_sitasi_sinta
Tipe Tabel	:	Multi table
Tipe Pemrograman	:	Bahasa PHP dan Library
Status Kebutuhan	:	Terpenuhi

### 5.3.1.3 Skor Kualitas Sitasi

Metode	:	Teknik <i>Grabbing</i>
Sasaran	:	Penyedia informasi dari database Google Scholar dan SINTA Ristek Dikti
	-	<a href="https://scholar.google.co.id/">https://scholar.google.co.id/</a>
	-	<a href="http://sinta1.ristekdikti.go.id/">http://sinta1.ristekdikti.go.id/</a>
Masukan	:	Halaman website pada poin sasaran
Keluaran	:	Hasil informasi dalam beberapa atribut
	-	nama_universitas
	-	jumlah_sitasi_scholar
	-	jumlah_i10_index_scholar
	-	jumlah_h_index_scholar
Jumlah Data	:	3471 data
Tabel Database	:	
	-	tampung_author_sinta
	-	tampung_author_sinta_2
	-	tapung_log_author_sinta
	-	jumlah_author_sinta
	-	jumlah_author_sinta_2
Tipe Tabel	:	Multi table
Tipe Pemrograman	:	Bahasa PHP dan Library
Status Kebutuhan	:	Terpenuhi

#### 5.3.1.4 Skor Jumlah Paper

Metode	: Teknik <i>Grabbing</i>
Sasaran Dikti	: Penyedia informasi dari database SINTA Ristek
	- <a href="http://sinta1.ristekdikti.go.id/">http://sinta1.ristekdikti.go.id/</a>
Masukan	: Halaman website pada poin sasaran
Keluaran	: Hasil informasi dalam beberapa atribut
	- nama_universitas
	- jumlah_paper_scholar
Jumlah Data	: 3471 data
Tabel Database	:
	- tampung_sitasi_sinta
	- tampung_sitasi_sinta_2
	- tapung_log_sitasi_sinta
	- tampung_sitasi_sinta_3
	- jumlah_sitasi_sinta
Tipe Tabel	: Multi table
Tipe Pemrograman	: Bahasa PHP dan Library
Status Kebutuhan	: Terpenuhi

#### 5.3.1.5 Skor Jumlah Artikel Jurnal

Metode	: Teknik <i>Grabbing</i>
Sasaran Dikti	: Penyedia informasi dari database SINTA Ristek
	- <a href="http://sinta1.ristekdikti.go.id/">http://sinta1.ristekdikti.go.id/</a>
Masukan	: Halaman website pada poin sasaran
Keluaran	: Hasil informasi dalam beberapa atribut
	- nama_universitas
	- jumlah_jurnal_artikel
Jumlah Data	: 3471 data
Tabel Database	:
	- tampung_sitasi_sinta
	- tampung_sitasi_sinta_2

	-	tapung_log_sitasi_sinta
	-	tampung_sitasi_sinta_3
	-	jumlah_sitasi_sinta
Tipe Tabel	:	Multi table
Tipe Pemrograman	:	Bahasa PHP dan Library
Status Kebutuhan	:	Terpenuhi

#### 5.3.1.6 Skor Jumlah Paper Conference

Metode	:	Teknik <i>Grabbing</i>
Sasaran	:	Penyedia informasi dari database SINTA Ristek Dikti
	-	<a href="http://sinta1.ristekdikti.go.id/">http://sinta1.ristekdikti.go.id/</a>
Masukan	:	Halaman website pada poin sasaran
Keluaran	:	Hasil informasi dalam beberapa atribut
	-	nama_universitas
	-	jumlah_paper_conference
Jumlah Data	:	3471 data
Tabel Database	:	
	-	tampung_sitasi_sinta
	-	tampung_sitasi_sinta_2
	-	tapung_log_sitasi_sinta
	-	tampung_sitasi_sinta_3
	-	jumlah_sitasi_sinta
Tipe Tabel	:	Multi table
Tipe Pemrograman	:	Bahasa PHP dan Library
Status Kebutuhan	:	Terpenuhi

#### 5.3.1.7 Skor Precense

Metode	:	Teknik <i>Grabbing</i>
Sasaran	:	Website mesin pencari Google (index Google)
	-	<a href="http://www.google.com/search">http://www.google.com/search</a>

Masukan : Domain website perguruan tinggi yang diaplikasikan menggunakan library dan indexing website pada poin sasaran

Keluaran : Hasil informasi dalam beberapa atribut

- url\_web
- domain\_search
- hasil\_search
- tanggal\_akses

Jumlah Data : 442 data

Tabel Database : precense

Tipe Tabel : Single table

Tipe Pemrograman : Bahasa PHP dan Library

Status Kebutuhan : Terpenuhi

#### 5.3.1.8 Skor Impact

Metode : Teknik *Grabbing*

Sasaran : Penyedia informasi analisa SEO menggunakan

MOZ Analytics

- <http://lsapi.seomoz.com/linkscape/url-metrics/>

Masukan : Domain website perguruan tinggi yang diaplikasikan menggunakan API dan Token user MOZ Analytics

Keluaran : Hasil informasi dalam beberapa atribut

- url\_web
- domain\_web
- page\_authority
- domain\_authority
- backlink\_eksternal
- tanggal\_akses

Jumlah Data : 447 data

Tabel Database : impact

Tipe Tabel : Single table

Tipe Pemrograman : Bahasa PHP dan Library

Status Kebutuhan : Terpenuhi



### 5.3.2 Skenario Pengambilan Data pada Tabel Database

Skenario pengambilan data pada database digunakan pada data yang bersifat pengolahan lokal (tanpa campur tangan koneksi internet) yang didapatkan dari pengolahan data online yang sudah disimpan dalam database maupun dari komputasi metode teks mining.

#### 5.3.2.1 Skor Reputasi Akademik

Metode : Analisa Database  
 Masukan : Hasil skor dari metode teks mining PLSA terhadap dokumen abstrak *paper* setiap perguruan tinggi  
 Keluaran : Hasil informasi dalam beberapa atribut

- nama\_universitas
- skor\_universitas
- bobot\_total
- jumlah\_dokumen
- tanggal\_akses

Jumlah Data : 122 data  
 Tabel Database : reputasi\_akademik  
 Tipe Tabel : Single table  
 Tipe Pemrograman : Bahasa Python dan Library  
 Status Kebutuhan : Terpenuhi

#### 5.3.2.2 Skor Jumlah Mahasiswa

Metode : Analisa Database  
 Masukan : Dokumen rekap jumlah mahasiswa dari database PDDIKTI  
 Keluaran : Hasil informasi dalam beberapa atribut

- kode\_universitas
- nama\_universitas
- jumlah\_tahun1\_semester1
- jumlah\_tahun1\_semester2

- jumlah\_tahun2\_semester1
- jumlah\_tahun2\_semester2
- status\_universitas

Jumlah Data : 3247 data

Tabel Database : jumlah\_mahasiswa

Tipe Tabel : Single table

Tipe Pemrograman : Bahasa Python dan Library

Status Kebutuhan : Terpenuhi

### 5.3.2.3 Skor Jumlah Tenaga Pengajar

Metode : Analisa Database

Masukan : Dokumen rekap jumlah tenaga pengajar dari database PDDIKTI

Keluaran : Hasil informasi dalam beberapa atribut

- kode\_universitas
- nama\_universitas
- asisten\_ahli\_10000
- asisten\_ahli\_15000
- lektor\_20000
- lektor\_30000
- lektor\_kepala\_40000
- lektor\_kepala\_55000
- lektor\_kepala\_70000
- profesor\_85000
- profesor\_105000
- tanpa\_jabatan
- dosen\_tetap
- dosen\_tidak\_tetap
- status\_universitas
- pendidikan\_terakhir

Jumlah Data : 3251 data

Tabel Database :

- jumlah\_tenaga\_pengajar\_jabatan\_akademik

- jumlah\_tenaga\_pengajar\_status\_dosen
- jumlah\_tenaga\_pengajar\_tetap
- jumlah\_tenaga\_pengajar\_tidak\_tetap

Tipe Tabel : Multi table

Tipe Pemrograman : Bahasa Python dan Library

Status Kebutuhan : Terpenuhi

#### 5.3.2.4 Skor Jumlah PhD

Metode : Analisa Database

Masukan : Dokumen rekap jumlah tenaga pengajar dari database PDDIKTI

Keluaran : Hasil informasi dalam beberapa atribut

- kode\_universitas
- nama\_universitas
- pendidikan\_terakhir\_S3

Jumlah Data : 3251 data

Tabel Database :

- jumlah\_tenaga\_pengajar\_tetap
- jumlah\_tenaga\_pengajar\_tidak\_tetap

Tipe Tabel : Multi table

Tipe Pemrograman : Bahasa Python dan Library

Status Kebutuhan : Terpenuhi

#### 5.3.2.5 Skor Transparency

Metode : Analisa Database

Masukan : Hasil query data jumlah sitasi Google Scholar pada tabel jumlah\_sitasi\_sinta

Keluaran : Hasil informasi dalam beberapa atribut

- nama\_universitas
- jumlah\_sitasi\_scholar

Jumlah Data : 3471 data

Tabel Database : jumlah\_sitasi\_sinta

Tipe Tabel : Single table  
 Tipe Pemrograman : Bahasa PHP dan Library  
 Status Kebutuhan : Terpenuhi

#### 5.3.2.6 Skor Excellence

Metode : Analisa Database  
 Masukan : Hasil query data jumlah paper Google Scholar pada tabel `jumlah_sitasi_sinta`  
 Keluaran : Hasil informasi dalam beberapa atribut
 

- `nama_universitas`
- `jumlah_paper_scholar`

 Jumlah Data : 3471 data  
 Tabel Database : `jumlah_sitasi_sinta`  
 Tipe Tabel : Single table  
 Tipe Pemrograman : Bahasa PHP dan Library  
 Status Kebutuhan : Terpenuhi

### 5.3.3 Skenario Pengujian Dataset Menggunakan Teks Mining Metode PLSA

#### 5.3.3.1 Tahap Preprocessing

Dataset dokumen abstrak *paper* yang digunakan berasal dari 16 perguruan tinggi teratas yang masuk dalam *Ground Truth QS World University Rankings 2016-2017*. Rincian jumlah kata asli maupun kata hasil reduksi dalam tahap preprocessing seperti pada Tabel 5.3-1.

**Tabel 5.3-1 Hasil Reduksi Kata Tahap Preprocessing**

No.	Nama Perguruan Tinggi	Jumlah Kata Asli	Jumlah Kata Hasil Preprocessing	Reduksi Kata
-----	-----------------------	------------------	---------------------------------	--------------

1.	Institut Pertanian Bogor	10539	6293	59,712%
2.	Institut Teknologi Bandung	9287	5675	61,107%
3.	Institut Teknologi Sepuluh Nopember	9630	6015	62,461%
4.	Universitas 17 Agustus 1945	9146	5580	61,01%
5.	Universitas Airlangga	9457	6009	63,54%
6.	Universitas Bina Nusantara	7236	4432	61,249%
7.	Universitas Brawijaya	8651	5469	63,218%
8.	Universitas Diponegoro	8397	5327	63,439%
9.	Universitas Gadjah Mada	10214	6279	61,474%
10.	Universitas Indonesia	8823	5645	63,981%
11.	Universitas Islam Indonesia	7897	4989	63,176%
12.	Universitas Mataram	7527	4561	60,595%
13.	Universitas Muhammadiyah Surakarta	10330	6117	59,216%
14.	Universitas Padjadjaran	7774	5008	64,42%
15.	Universitas Pelita Harapan	7204	4542	63,048%

16.	Universitas Sebelas Maret	7807	5244	67,17%
Rata-rata				62,426%

Berdasarkan Tabel 5.3-1 didapatkan rata-rata hasil reduksi kata dari semua dataset sebesar 62,426%, yang menunjukkan tahap preprocessing mampu menghapus karakter kata yang tidak penting sebesar 37,574%.

Jumlah duplikasi kata yang tinggi dalam sebuah dokumen menjadi tingkat kepentingan sebuah kata menjadi kurang bagus. Tabel 5.3-2 menampilkan jumlah tingkat duplikasi kata dalam dokumen abstrak *paper* akademisi di 16 perguruan tinggi.

**Tabel 5.3-2 Tingkat Duplikasi Kata**

No	Nama Perguruan Tinggi	Jumlah Kata Hasil Preprocessing	Jumlah Kata (duplikasi)	Jumlah Kata (single)	Duplikasi Kata (%)
1.	Institut Pertanian Bogor	6293	4469	1824	71,051%
2.	Institut Teknologi Bandung	5675	3929	1746	69,233%
3.	Institut Teknologi Sepuluh Nopember	6015	3879	2136	64,489%
4.	Universitas 17 Agustus 1945	5580	3576	2004	64,086%
5.	Universitas Airlangga	6009	4234	1775	70,461%
6.	Universitas Bina Nusantara	4432	3004	1428	67,78%
7.	Universitas Brawijaya	5469	3581	1888	65,478%

8.	Universitas Diponegoro	5327	3519	1808	66,06%
9.	Universitas Gadjah Mada	6279	3979	2300	63,37%
10.	Universitas Indonesia	5645	3641	2004	64,5%
11.	Universitas Islam Indonesia	4989	3036	1953	60,854%
12.	Universitas Mataram	4561	3253	1308	71,322%
13.	Universitas Muhammadi yah Surakarta	6117	4123	1994	67,402%
14.	Universitas Padjadjaran	5008	3203	1805	63,958%
15.	Universitas Pelita Harapan	4542	2994	1548	65,918%
16.	Universitas Sebelas Maret	5244	3806	1538	71,22%
Rata-rata					66,697%

Berdasarkan Tabel 5.3-2 didapatkan rata-rata duplikasi kata dalam dokumen hasil preprocessing sebesar 66,697% yang menunjukkan duplikasi kata sangat besar.

### 5.3.3.2 Tahap Memperkaya Corpus dengan POS Tagging

Corpus TKT yang didapatkan dari proses Breakdown Corpus Taxonomy Bloom akan dikembangkan lagi agar menghasilkan keyword yang lebih banyak. Pengembangan ini menggunakan metode POS Tagging. Rincian jumlah kata penambahan dengan menggunakan POS Tagging terhadap default pada corpus TKT seperti pada Tabel 5.3-3.

**Tabel 5.3-3 Rincian Jumlah Penambahan Kata dengan POS  
Tagging Setiap Perguruan Tinggi**

No.	Nama Perguruan Tinggi	POS Tagging Akuisisi Ke-					
		1	2	3	4	8	10
1.	Institut Pertanian Bogor	201	329	457	585	1097	1353
2.	Institut Teknologi Bandung	149	241	333	425	793	977
3.	Institut Teknologi Sepuluh Nopember	215	349	483	617	1086	1220
4.	Universitas 17 Agustus 1945	226	386	510	652	1220	1291
5.	Universitas Airlangga	224	364	504	644	1134	1274
6.	Universitas Bina Nusantara	225	369	513	657	1233	1521
7.	Universitas Brawijaya	189	309	429	549	1029	1269
8.	Universitas Diponegoro	177	287	397	507	947	1167
9.	Universitas Gadjah Mada	261	343	425	507	835	999
10.	Universitas Indonesia	227	367	507	647	1207	1487
11.	Universitas Islam Indonesia	226	366	506	646	1206	1486
12.	Universitas Mataram	126	206	286	366	686	846
13.	Universitas Muhammadiyah	212	346	480	614	1150	1418



	yah Surakarta						
14.	Universitas Padjadjaran	234	380	526	672	1256	1548
15.	Universitas Pelita Harapan	180	292	404	516	964	1188
16.	Universitas Sebelas Maret	159	259	359	459	659	809
Total Keseluruhan		3241	5209	7141	9091	16554	19917
Rata-rata Penambahan		3,7%	6,0%	8,2%	10,5%	19,1%	23,1%

Berdasarkan Tabel 5.3-3 didapatkan rata-rata penambahan kata baru menggunakan teknik POS Tagging dengan variasi akusisi yang berbeda-beda. Dapat dilihat rata-rata tertinggi adalah 23,1% untuk variasi akusisi 10.

Semakin besar akusisi maka presentase penambahan juga semakin tinggi, akan tetapi perlu dilihat juga tingkat kata-kata duplikasi yang dihasilkan. Oleh karena itu tidak bisa menjadi jaminan jika prosentase jumlah penambahan kata tinggi akan menghasilkan corpus yang besar dan benar, perlu adanya proses lebih lanjut untuk menganalisa dan mereduksi kata duplikasi. Jumlah breakdown kata dari hasil POS Tagging per dataset paper pada corpus TKT seperti pada Tabel 5.3-4.

**Tabel 5.3-4 Rincian Jumlah Breakdown Hasil POS Tagging pada Corpus TKT**

No.	Level	POS Tagging Akusisi Ke-					
		1	2	3	4	8	10
1.	TKT 1	520	927	1326	1698	3235	3951
2.	TKT 2	358	552	766	971	1722	2091
3.	TKT 3	507	816	1124	1442	2629	1260
4.	TKT 4	462	738	1014	1284	2333	2818
5.	TKT 5	321	528	728	940	1731	2081
6.	TKT 6	421	637	876	1098	1905	1168

7.	TKT 7	236	368	515	650	1211	1422
8.	TKT 8	158	216	275	343	542	628
9.	TKT 9	258	409	517	665	1246	1498
Total Keseluruhan		3241	5209	7141	9091	16554	19917

Hasil breakdown pada Tabel 5.3-4 dihasilkan jumlah kata untuk corpus final TKT, corpus ini akan menjadi dasar untuk proses metode teks mining PLSA pada percobaan menggunakan teknik POS Tagging. Rincian corpus final TKT seperti pada Tabel 5.3-5.

**Tabel 5.3-5 Corpus TKT Final Berdasarkan Percobaan POS Tagging**

No.	Level	Jumlah kata default	Corpus final dengan POS Tagging Akuisisi Ke-					
			1	2	3	4	8	10
1.	TKT 1	31	379	610	789	917	1313	1481
2.	TKT 2	24	233	328	420	506	735	820
3.	TKT 3	32	321	489	611	741	1032	1123
4.	TKT 4	15	275	436	561	668	1010	1139
5.	TKT 5	31	212	332	425	504	741	786
6.	TKT 6	24	228	324	407	463	587	618
7.	TKT 7	26	156	224	295	347	527	584
8.	TKT 8	30	90	122	159	193	288	323
9.	TKT 9	34	169	259	300	345	517	581
Total Keseluruhan		256	2063	3124	3967	4684	6750	7465

Jumlah kata pada masing-masing level corpus TKT final berdasarkan penambahan POS Tagging dijabarkan pada poin-poin dibawah ini:

1. Akuisisi POS Tagging 1 (-1, +1)

**Tabel 5.3-6 Analisa Kata Hasil Akuisisi POS Tagging 1 (-1, +1)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	

1.	IPB	114	67	93	81	66	78	46	15	44	1220
2.	ITB	106	72	93	70	62	58	41	11	42	1191
3.	ITS	120	63	107	83	78	84	53	13	55	1480
4.	UNTAG	101	72	89	82	69	78	39	16	54	1404
5.	UNAIR	109	65	95	75	56	87	41	13	59	1175
6.	BINUS	118	58	92	84	56	74	37	14	42	853
7.	UB	117	72	111	92	67	82	35	9	47	1256
8.	UNDIP	106	59	102	68	60	75	37	12	40	1249
9.	UGM	139	75	125	105	87	85	47	18	55	1564
10.	UI	123	60	121	96	65	85	51	18	49	1336
11.	UIN	109	75	94	84	59	76	42	12	58	1344
12.	UNRAM	93	46	79	58	55	59	30	16	34	838
13.	UMS	112	64	99	94	69	90	48	18	49	1351
14.	UNPAD	123	75	97	69	57	80	52	11	43	1198
15.	UPH	98	64	99	65	67	73	35	15	38	994
16.	UNS	86	59	95	78	74	65	38	7	32	1004

Berdasarkan POS Tagging akuisisi 1 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 33,30%.

## 2. Akuisisi POS Tagging 2 (-2, +2)

**Tabel 5.3-7 Analisa Kata Hasil Akuisisi POS Tagging 2 (-2, +2)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	
1.	IPB	180	78	134	117	97	100	62	15	65	976
2.	ITB	173	90	140	100	85	78	57	20	56	947
3.	ITS	200	82	158	143	117	118	68	22	74	1154
4.	UNTAG	164	98	131	121	90	107	54	20	79	1140
5.	UNAIR	178	87	132	112	84	114	51	22	86	909
6.	BINUS	167	82	139	113	85	104	48	14	64	612
7.	UB	175	101	154	121	104	110	50	15	72	986
8.	UNDIP	176	79	148	104	93	99	51	21	60	977
9.	UGM	210	94	179	151	120	116	64	21	76	1269
10.	UI	196	86	170	113	102	114	71	22	65	1065
11.	UIN	175	94	143	118	86	102	57	21	80	1077

12.	UNRAM	143	58	111	87	86	74	46	16	51	636
13.	UMS	184	89	152	135	101	122	68	25	69	1049
14.	UNPAD	200	93	136	99	88	106	68	16	62	937
15.	UPH	156	81	137	91	94	95	53	18	59	764
16.	UNS	138	77	138	118	103	86	49	12	54	763

Berdasarkan POS Tagging akuisisi 2 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 47,84%.

### 3. Akuisisi POS Tagging 3 (-3, +3)

**Tabel 5.3-8 Analisa Kata Hasil Akuisisi POS Tagging 3 (-3, +3)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	
1.	IPB	219	97	158	145	111	117	70	16	78	813
2.	ITB	219	100	157	123	101	94	67	24	61	800
3.	ITS	248	102	178	174	143	145	78	24	81	963
4.	UNTAG	208	114	155	150	109	131	66	24	80	967
5.	UNAIR	221	106	154	134	93	135	72	24	94	742
6.	BINUS	198	98	162	135	99	130	55	17	73	461
7.	UB	220	118	180	158	123	137	59	17	77	799
8.	UNDIP	215	94	171	131	107	115	66	21	69	819
9.	UGM	262	115	193	187	141	141	74	21	80	1086
10.	UI	246	98	197	182	117	139	80	27	73	845
11.	UIN	219	122	162	150	106	118	64	23	87	902
12.	UNRAM	168	70	127	104	95	85	56	25	53	525
13.	UMS	238	110	170	161	119	150	79	30	87	850
14.	UNPAD	238	117	172	120	104	128	83	20	66	757
15.	UPH	202	104	157	110	105	115	65	23	65	602
16.	UNS	179	86	156	135	123	101	57	15	62	624

Berdasarkan POS Tagging akuisisi 3 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 57,17%.

### 4. Akuisisi POS Tagging 4 (-4, +4)

**Tabel 5.3-9 Analisa Kata Hasil Akuisisi POS Tagging 4 (-4, +4)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	
1.	IPB	246	107	180	163	126	131	74	19	85	693
2.	ITB	249	119	170	129	113	106	73	27	66	694
3.	ITS	282	113	203	193	155	155	82	27	90	836
4.	UNTAG	222	129	178	167	122	145	71	28	93	849
5.	UNAIR	232	121	169	157	104	147	68	33	102	642
6.	BINUS	219	108	179	151	107	143	60	18	79	364
7.	UB	239	133	203	171	145	152	66	23	86	670
8.	UNDIP	243	111	193	153	131	129	64	23	71	690
9.	UGM	294	128	208	215	166	157	82	30	84	936
10.	UI	269	119	227	195	139	155	82	29	82	707
11.	UIN	242	137	185	162	118	134	76	29	89	781
12.	UNRAM	186	76	149	116	113	93	61	26	51	437
13.	UMS	262	131	204	180	138	162	88	33	87	709
14.	UNPAD	269	129	198	144	112	144	86	21	71	631
15.	UPH	219	114	177	118	127	126	69	25	68	505
16.	UNS	197	99	176	148	138	107	60	18	70	525

Berdasarkan POS Tagging akuisisi 4 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 63,77%.

#### 5. Akuisisi POS Tagging 8 (-8, +8)

**Tabel 5.3-10 Analisa Kata Hasil Akuisisi POS Tagging 8 (-8, +8)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	
1.	IPB	300	132	217	212	155	154	93	24	105	432
2.	ITB	324	143	200	153	134	120	98	20	72	482
3.	ITS	358	146	234	231	200	166	104	34	105	558
4.	UNTAG	276	159	205	197	143	167	100	38	108	611
5.	UNAIR	285	144	214	200	122	164	83	32	109	422
6.	BINUS	269	127	216	182	123	151	69	20	88	183

7.	UB	302	160	233	210	171	172	84	20	96	440
8.	UNDIP	287	131	210	191	158	142	93	29	92	475
9.	UGM	365	163	254	259	193	177	101	32	95	661
10.	UI	339	150	268	244	166	175	114	32	96	420
11.	UII	311	182	226	203	143	156	92	32	111	497
12.	UNRAM	228	98	169	142	137	103	77	29	61	264
13.	UMS	328	159	240	223	164	186	108	37	105	444
14.	UNPAD	348	167	235	177	135	170	106	25	85	357
15.	UPH	278	143	203	152	128	143	86	26	78	311
16.	UNS	241	116	202	178	168	119	68	19	74	353

Berdasarkan POS Tagging akuisisi 8 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 76,59%.

#### 6. Akuisisi POS Tagging 10 (-10, +10)

**Tabel 5.3-11 Analisa Kata Hasil Akuisisi POS Tagging 10 (-10, +10)**

No.	Nama	Level TKT									Not in TKT
		1	2	3	4	5	6	7	8	9	
1.	IPB	318	135	234	247	161	156	95	27	107	344
2.	ITB	355	150	215	157	131	122	106	21	78	411
3.	ITS	375	151	242	241	198	169	113	40	105	502
4.	UNTAG	288	169	214	208	145	170	101	39	106	564
5.	UNAIR	303	161	211	202	130	164	83	32	112	377
6.	BINUS	279	132	221	181	126	156	73	20	88	152
7.	UB	320	159	243	226	181	175	83	26	101	374
8.	UNDIP	298	138	222	200	170	147	97	26	85	425
9.	UGM	388	169	260	274	203	184	104	29	105	584
10.	UI	361	158	276	260	169	180	121	31	104	344
11.	UII	334	187	233	215	149	157	97	38	119	424
12.	UNRAM	235	104	181	159	137	103	79	29	60	221
13.	UMS	345	166	256	244	170	188	109	40	101	375
14.	UNPAD	357	173	255	191	143	175	110	29	91	281
15.	UPH	296	154	208	157	134	146	87	28	88	250
16.	UNS	260	126	208	183	175	120	74	15	81	296

Berdasarkan POS Tagging akuisisi 10 didapatkan rata-rata keberadaan kata dari dataset dokumen abstrak paper sebesar 79,98%.

Dalam Tugas Akhir ini disimpulkan menggunakan akuisisi kata pada POS Tagger 1 untuk mengurangi waktu komputasi. Berdasarkan hasil analisa terdapat kesamaan pola antar pengambilan kata menggunakan akuisisi 1 hingga 10, yang berbeda adalah jumlah kata saja. Akan tetapi pola jumlah kata per corpus TKT tetap sama.

#### 5.3.3.3 Memperkaya Corpus TKT dengan Sinonim Word

Selain menggunakan percobaan memperkaya corpus TKT dengan teknik POS Tagging, Tugas Akhir ini juga menggunakan pengembangan dari segi sinonim word. Pembahasan lebih dalam telah dijelaskan dalam bab 5.2.3.3. penggunaan sinonim word sebagai salah satu alternative pengembangan corpus TKT dari segi corpus, bukan dari segi dokumen.

#### 5.3.3.4 Tahap Menentukan Asumsi Label Topik Awal

Menentukan asumsi label topik awal menggunakan Term Frequency (TF) menjadi tahap yang paling berperan dalam metode teks mining dalam Tugas Akhir. Beberapa percobaan untuk Menentukan asumsi label topik awal seperti berikut:

##### 1. Percobaan 1

- Dataset : Dokumen hasil preprocessing perguruan tinggi
- Corpus : Corpus TKT final dengan POS Tagging akuisisi 1
- Proses similarity : Term Frequency
- Jumlah label : 3 label teratas

**Tabel 5.3-12 Label Dokumen Hasil Percobaan Teknik POS Tagging Akuisisi 1**

No.	Nama	Asumsi Label Awal
-----	------	-------------------

1.	IPB	T6, T7, T5
2.	ITB	T6, T7, T5
3.	ITS	T6, T7, T5
4.	UNTAG	T6, T7, T5
5.	UNAIR	T6, T7, T5
6.	BINUS	T6, T7, T5
7.	UB	T6, T7, T5
8.	UNDIP	T6, T7, T5
9.	UGM	T6, T7, T5
10.	UI	T6, T7, T5
11.	UIN	T6, T7, T5
12.	UNRAM	T6, T7, T5
13.	UMS	T6, T7, T5
14.	UNPAD	T6, T7, T5
15.	UPH	T6, T7, T5
16.	UNS	T6, T7, T5

## 2. Percobaan 2

- Dataset : Dokumen hasil preprocessing perguruan tinggi
- Corpus : Corpus TKT final dengan sinonim word level all (semua sinonim digunakan)
- Proses similarity : Term Frequency
- Jumlah label : 2 label teratas

**Tabel 5.3-13 Label Dokumen Hasil Percobaan Sinonim Word Level All**

No.	Nama	Asumsi Label Awal
1.	IPB	T5, T7, T9
2.	ITB	T5, T7, T9
3.	ITS	T5, T7, T9
4.	UNTAG	T5, T7, T9
5.	UNAIR	T5, T7, T9
6.	BINUS	T5, T7, T9
7.	UB	T5, T7, T9



8.	UNDIP	T5, T7, T9
9.	UGM	T5, T7, T9
10.	UI	T5, T7, T9
11.	UII	T5, T7, T9
12.	UNRAM	T5, T7, T9
13.	UMS	T5, T7, T9
14.	UNPAD	T5, T7, T9
15.	UPH	T5, T7, T9
16.	UNS	T5, T7, T9

### 5.3.3.5 Tahap Generatif Model Supervised PLSA

Berdasarkan percobaan pada tahap preprocessing, memperkaya corpus TKT, dan penentuan asumsi label topik awal dilakukan beberapa percobaan pada tahap *Supervised* PLSA yang nantinya akan dibandingkan hasil akhir yang paling bagus didapatkan dari langkah apa saja. Berikut adalah hasil generative PLSA dari beberapa percobaan:

#### 1. Percobaan 1

Berdasarkan percobaan 1 dalam tahap memperkaya corpus didapatkan hasil probabilitas topik terhadap dokumen seperti pada Tabel 5.3-18.

**Tabel 5.3-14 Probabilitas Label dengan Dokumen dalam Percobaan 1**

No.	Nama	Level TKT								
		1	2	3	4	5	6	7	8	9
1.	IPB	0	0	0	0	0.148922	0.148922	0.148922	0	0
2.	ITB	0	0	0	0	0.148928	0.148928	0.148928	0	0
3.	ITS	0	0	0	0	0.148935	0.148935	0.148935	0	0
4.	UNTAG	0	0	0	0	0.148949	0.148949	0.148949	0	0
5.	UNAIR	0	0	0	0	0.14894	0.14894	0.14894	0	0
6.	BINUS	0	0	0	0	0.148937	0.148937	0.148937	0	0
7.	UB	0	0	0	0	0.148936	0.148936	0.148936	0	0
8.	UNDIP	0	0	0	0	0.148936	0.148936	0.148936	0	0
9.	UGM	0	0	0	0	0.148937	0.148937	0.148937	0	0

10.	UI	0	0	0	0	0.148935	0.148935	0.148935	0	0
11.	UIN	0	0	0	0	0.148946	0.148946	0.148946	0	0
12.	UNRAM	0	0	0	0	0.148929	0.148929	0.148929	0	0
13.	UMS	0	0	0	0	0.148936	0.148936	0.148936	0	0
14.	UNPAD	0	0	0	0	0.148942	0.148942	0.148942	0	0
15.	UPH	0	0	0	0	0.148938	0.148938	0.148938	0	0
16.	UNS	0	0	0	0	0.148933	0.148933	0.148933	0	0

## 2. Percobaan 2

Berdasarkan percobaan 2 dalam tahap memperkaya corpus didapatkan hasil probabilitas topik terhadap dokumen seperti pada Tabel 5.3-19.

**Tabel 5.3-15 Probabilitas Label dengan Dokumen dalam Percobaan 2**

No.	Nama	Level TKT								
		1	2	3	4	5	6	7	8	9
1.	IPB	0	0	0	0	0.148922	0	0.148922	0	0.170209
2.	ITB	0	0	0	0	0.148928	0	0.148928	0	0.17023
3.	ITS	0	0	0	0	0.148935	0	0.148935	0	0.170214
4.	UNTAG	0	0	0	0	0.148949	0	0.148949	0	0.170195
5.	UNAIR	0	0	0	0	0.14894	0	0.14894	0	0.170204
6.	BINUS	0	0	0	0	0.148937	0	0.148937	0	0.170212
7.	UB	0	0	0	0	0.148936	0	0.148936	0	0.170213
8.	UNDIP	0	0	0	0	0.148936	0	0.148936	0	0.170214
9.	UGM	0	0	0	0	0.148937	0	0.148937	0	0.170213
10.	UI	0	0	0	0	0.148935	0	0.148935	0	0.170218
11.	UIN	0	0	0	0	0.148946	0	0.148946	0	0.170196
12.	UNRAM	0	0	0	0	0.148929	0	0.148929	0	0.170237
13.	UMS	0	0	0	0	0.148936	0	0.148936	0	0.170214
14.	UNPAD	0	0	0	0	0.148942	0	0.148942	0	0.170206
15.	UPH	0	0	0	0	0.148938	0	0.148938	0	0.170218
16.	UNS	0	0	0	0	0.148933	0	0.148933	0	0.170224

### 5.3.3.6 Tahap Pembobotan Final Pemeringkatan

Hasil keluaran dari tahap Generatif Model PLSA akan dihitung menggunakan bobot yang sudah ditentukan. Dengan rincian setiap percobaan sebagai berikut.

#### 1. Percobaan 1

**Tabel 5.3-20 Topik Akhir Dokumen Menggunakan Percobaan Teknik POS Tagging Akuisisi 1**

No.	Nama	Label Topik AKhir
1.	IPB	T6
2.	ITB	T6
3.	ITS	T5
4.	UNTAG	T6
5.	UNAIR	T7
6.	BINUS	T6
7.	UB	T5
8.	UNDIP	T6
9.	UGM	T7
10.	UI	T7
11.	UIN	T5
12.	UNRAM	T5
13.	UMS	T6
14.	UNPAD	T7
15.	UPH	T5
16.	UNS	T5

#### 2. Percobaan 2

**Tabel 5.3-21 Topik Akhir Dokumen Menggunakan Percobaan Sinonim Word Level All**

No.	Nama	Label Topik AKhir
1.	IPB	T7
2.	ITB	T7
3.	ITS	T5

4.	UNTAG	T7
5.	UNAIR	T7
6.	BINUS	T7
7.	UB	T7
8.	UNDIP	T7
9.	UGM	T7
10.	UI	T7
11.	UIN	T5
12.	UNRAM	T5
13.	UMS	T7
14.	UNPAD	T7
15.	UPH	T7
16.	UNS	T7

#### 5.3.3.7 Menghitung Distance Hasil Sistem dengan Ground Truth QS World University Rankings 2016-2017

Distance dalam hal ini adalah beda peringkat hasil sistem dengan peringkat yang ada pada Ground Truth. Semakin kecil distance maka semakin bagus peringkat sistem yang dihasilkan. Berikut adalah analisa distance dari 2 percobaan yang sudah dilakukan diatas.

##### 1. Percobaan 1

**Tabel 5.3-16 Perbandingan Peringkat Ground Truth QS World University Rankings 2016-2017 dengan Hasil Percobaan 1**

Nama	Peringkat Ground Truth	Peringkat Sistem	Beda	Flag Toleransi ( $\leq 3$ )
IPB	5	6	1	1
ITB	1	1	0	1
ITS	7	9	2	1
UB	9	8	1	1
UGM	3	3	0	1
UI	2	2	0	1

UMS	8	4	4	0
UNAIR	4	5	1	1
UNDIP	6	7	1	1
<b>Total Distance</b>			<b>10</b>	<b>8</b>

## 2. Percobaan 2

**Tabel 5.3-17 Perbandingan Peringkat Ground Truth QS World University Rankings 2016-2017 dengan Hasil Percobaan 2**

Nama	Peringkat Ground Truth	Peringkat Sistem	Beda	Flag Toleransi ( $\leq 3$ )
IPB	5	9	4	0
ITB	1	1	0	1
ITS	7	7	0	1
UB	9	6	3	1
UGM	3	3	0	1
UI	2	2	0	1
UMS	8	4	4	0
UNAIR	4	8	4	0
UNDIP	6	5	1	1
<b>Total Distance</b>			<b>16</b>	<b>6</b>

### 5.3.3.8 Perbandingan Hasil Analisa Masing-masing Percobaan dengan Ground Truth QS World University Rankings 2016-2017

**Tabel 5.3-18 Hasil Analisa Semua Percobaan dengan Ground Truth QS World University Rankings 2016-2017**

Lingkup	Indikator Analisa	Percobaan 1	Percobaan 2
Indikator Reputasi Akademik	Distance (beda)	10	16
	Toleransi	8/9	6/9
	Prosentase toleransi	88%	66%
	Similaritay Pearson Correlation	80 %	51%
	Analisa Uji Table T (kepercayaan 95%)	h0 diterima	h0 diterima
	Analisa Uji Table T (kepercayaan 60%)	h0 diterima	h0 diterima
Sistem Keseluruhan	Distance (beda)	13	41
	Toleransi	8/9	2/9
	Prosentase toleransi	88,89%	22%
	Similaritay Pearson Correlation	73,78%	50,40%

### 5.3.3.9 Perbandingan Hasil Analisa Masing-masing Percobaan dengan Ground Truth QS World University Rankings 2017-2018

**Tabel 5.3-19 Hasil Analisa Semua Percobaan dengan Ground Truth QS World University Rankings 2017-2018**

Lingkup	Indikator Analisa	Percobaan 1	Percobaan 2
Indikator Reputasi Akademik	Distance (beda)	10	16
	Toleransi	8/9	6/9
	Prosentase toleransi	88%	66%
	Similaritay Pearson Correlation	80 %	51%
	Analisa Uji Table T (kepercayaan 95%)	h0 diterima	h0 diterima
	Analisa Uji Table T (kepercayaan 60%)	h0 diterima	h0 diterima
Sistem Keseluruhan	Distance (beda)	16	31
	Toleransi	8/9	3/9
	Prosentase toleransi	88,89%	33,33%
	Similaritay Pearson Correlation	78,33%	55,81%

### 5.3.4 Skenario Pengujian Sitem Berdasarkan Hasil Pembobotan Akhir Sesuai *Ground Truth QS Worlds University Rankings*

#### 5.3.4.1 Melihat Peringkat Perguruan Tinggi Berdasarkan Tipe Pemeringkatan QS Metode PLSA

- Aktor : pengguna aplikasi website
- Langkah :
  1. Pengguna membuka aplikasi website perangkat lunak.
  2. Menuju halaman “Lihat Ranking” dengan scroll page website kebawah.
  3. Mengisi bobot sesuai tipe pemeringkatan PLSA, bobot bisa menggunakan konfigurasi default maupun dapat dirubah sesuai kebutuhan seperti Gambar 5.3.4.1.1.
  4. Pilih tombol “Cek QS PLSA” seperti Gambar 5.3.4.1.2.
  5. Sistem kan menampilkan hasil pemeringkatan sesuai tipe QS PLSA seperti Gambar 5.3-1.

**IDN**University  
Ranking
Metode Penilaian   Region   Lihat Ranking   Cari Disini

### Atur Bobot Pemeringkatan

Pembobotan Pemeringkatan									
Indikator	QS General	QS Asia	Webometrics	4ICU	DIKTI	QS (L-LDA)	QS (PLSA)	QS (Helmholtz)	QS (Adaboost)
Reputasi Akademik	<input type="text" value="40"/>	<input type="text" value="30"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="26"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Reputasi Employer	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="14"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Mahasiswa Aktif	<input type="text" value="20"/>	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Situs Akademisi	<input type="text" value="20"/>	<input type="text" value="10"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="15"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Tenaga Pengajar	<input type="text" value="5"/>	<input type="text" value="2.5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Kualitas Situs (F-index & I10-index)	<input type="text" value="5"/>	<input type="text" value="2.5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="3"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
Jumlah Phd (S3)	<input type="text" value="0"/>	<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="20"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>

**Gambar 5.3-1 Antarmuka Pembobotan Peringkat**



Jumlah Total	100	100	100	100	100	100	100	100	100
	Cek QS-General	Cek QS-Asia	Cek Webometrics	Cek 4ICU	Cek DIKT	Cek QS L-LDA	Cek QS PLSA	Cek QS Helmholtz	Cek QS Adaboost

Gambar 5.3-2 Antarmuka Tombol Cek Pemeringkatan

No.	Nama Perguruan Tinggi	Skor	Kota	Provinsi
1.	Universitas Indonesia	0.8085	Depok	Jawa Barat
2.	Institut Teknologi Bandung	0.7978	Bandung	Jawa Barat
3.	Universitas Gadjah Mada	0.7952	Yogyakarta	Di Yogyakarta
4.	Institut Pertanian Bogor	0.6802	Bogor	Jawa Barat
5.	Universitas Airlangga	0.6178	Surabaya	Jawa Timur
6.	Universitas Brawijaya	0.6117	Malang	Jawa Timur
7.	Universitas Diponegoro	0.6014	Semarang	Jawa Tengah
8.	Universitas Hasanuddin	0.5956	Makassar	Sulawesi Selatan
9.	Universitas Bina Nusantara	0.5939	Jakarta	DKI Jakarta
10.	Universitas Padjadjaran	0.575	Bandung	Jawa Barat
11.	Institut Teknologi Sepuluh Nopember	0.5566	Surabaya	Jawa Timur
12.	Universitas Udayana	0.5554	Badung	Bali

Gambar 5.3-3 Antarmuka Hasil Pemeringkatan

Berdasarkan skenario pengujian kinerja sistem terhadap Ground Truth *QS World University Ranking* didapatkan nilai similarity peringkat sebesar sebesar 78.33%.

5.4 Evaluasi

Rangkuman mengenai hasil pengujian fungsionalitas dapat dilihat pada Tabel 5.4.1. Berdasarkan data pada tabel tersebut, semua skenario pengujian berhasil dan program berjalan dengan baik. Sehingga bisa ditarik disimpulkan bahwa fungsionalitas dari program telah bisa bekerja sesuai dengan yang diharapkan.

Tabel 5.4-1 Rangkuman Hasil Pengujian

ID	Nama	Hasil Uji Coba
UJ-UC-001	Memasukkan presentase bobot kriteria penilaian pemeringkatan	Berhasil
UJ-UC-002	Meihat peringkat perguruan tinggi berdasar tipe pemeringkatan	Berhasil

UJ-UC-003	Mencari perguruan tinggi berdasarkan potongan nama	Berhasil
UJ-UC-004	Mencari perguruan tinggi berdasarkan region wilayah provinsi	Berhasil
UJ-UC-005	Melihat detail informasi utama perguruan tinggi	Berhasil

## BAB VI KESIMPULAN DAN SARAN

Pada bab ini akan diberikan kesimpulan yang diambil selama pengerjaan Tugas Akhir serta saran-saran tentang pengembangan yang dapat dilakukan terhadap Tugas Akhir ini di masa yang akan datang.

### 6.1 Kesimpulan

Dari hasil pengamatan selama proses perancangan, implementasi, dan pengujian perangkat lunak yang dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Metode teks minning menggunakan *Probabilistic Latent Semantic Analysis* dapat menggantikan kriteria reputasi akademik yang sebelumnya bersifat manual dengan menggunakan percobaan 1 yaitu menggunakan POST tagging dengan rincian kinerja sebagai berikut, distance peringkat sebesar 10, peringkat dalam toleransi sebesar 88,88%, dan similarity pearson corelation sebesar 80%.
2. Kinerja sistem secara keseluruhan dapat menyamai gorund truth QS World University Ranking 2016-2017, dengan menggunakan percobaan 1 yaitu menggunakan POST tagging dengan rincian sebagai berikut, distance peringkat sebesar 16, peringkat dalam toleransi sebesar 88,88%, dan similarity pearson correlation sebesar 78,33%,.
3. Metode teks minning dapat dikembangkan lagi menggunakan teknik POS Tagging untuk memperkaya corpus TKT yang digunakan sebagai acuan kelas topik final
4. Memperkaya corpus dapat menghasilkan hasil yang baik jika proses breakdown dan pengembangannya tepat. Perlu adanya analisa manual terhadap hasil memperkaya corpus dengan POS Tagging agar mendapatkan kelas kata yang tepat dan seimbang.
5. Kombinasi kriteria pemeringkatan yang bersifat kualitatif dan kuantitatif dapat menghasilkan hasil pemeringkatan yang mendekati Ground Truth sistem yaitu QS University Rankings 2016-2017.

6. Jumlah topik yang didefinisikan sebagai output dari algoritma PLSA sangat dipengaruhi oleh topik yang didefinisikan. Apabila beberapa topik yang didefinisikan merupakan topik yang cukup beririsan maka nilai akurasi dari variasi jumlah topik tidak akan jauh berbeda.
7. Penggunaan data berbasis SINTA dalam Tugas Akhir ini kurang valid karena data yang digunakan tidak diperbaharui secara berkala oleh pihak SINTA

## 6.2 Saran

Berikut merupakan beberapa saran untuk pengembangan sistem di masa yang akan datang. Saran-saran ini didasarkan pada hasil perancangan, implementasi dan pengujian yang telah dilakukan.

1. Melakukan penambahan kriteria pemeringkatan dapat dikondisikan dengan kebutuhan, untuk menghasilkan keragaman kriteria pemeringkatan dan tidak bertumpu pada satu kriteria utama saja.
2. Memberikan batas iterasi untuk proses perulangan penghitungan likelihood dengan cara menetapkan jumlah iterasi yang optimum ataupun dengan menetapkan nilai minimum perubahan likelihood yang dinilai signifikan
3. Menambahkan atau memperbaiki metode untuk memperkaya corpus dengan POS Tagging, berguna untuk menghasilkan corpus yang lebih efisien dan lengkap.
4. Perlu dilakukan analisa manual terhadap dokumen corpus hasil POS Tagging untuk memastikan kelas kata yang benar.
5. Pengembangan perangkat lunak kedepannya untuk data jumlah sitasi disarankan untuk menggunakan data dari Scopus karena data yang disediakan lengkap dan diperbaharui secara berkala

## DAFTAR PUSTAKA

- [1] K.S. Reddy, En Xie, Qingqing Tang, "Higher education, high-impact research, and world university rankings: A Case of India and comparison with China," *Pacific Science Review B: Humanities and Social Sciences* 2, pp. 1-21, 2016.
- [2] Tatiana Sidorenko, Tatiana Gorbatoeva, "Efficiency of Russian education through the scale of," *Procedia - Social and Behavioral Sciences* 166, pp. 464-467, 2015.
- [3] Mu-Hsuan Huang, "Opening The Black Box of QS World University Rankings," *Research Evaluation* 21, pp. 71-78, 2012.
- [4] QS World University Rankings, "QS Top University," 1 January 2016.[Online].Available:<http://www.topuniversities.com/university-rankings/asian-university-rankings/2016>. [Diakses 10 December 2016].
- [5] Kementerian, Riset Teknologi dan Pendidikan Tinggi, "Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 42 Tahun 2016 Tentang Pengukuran dan Penetapan Tingkat Kesiapterapan Teknologi," Kementerian Riset, Teknologi dan Pendidikan Tinggi, Jakarta, 2016.
- [6] Jin Wang, Ping Liu, Mary F.H. She, Abbas Kouzani, Saeid Nahavandi, "Supervised learning probabilistic Latent Semantic Analysis," 2012.
- [7] Wikipedia, "Wikipedia College and University Rankings," 14 December 2016. [Online]. Available: [https://en.wikipedia.org/wiki/College\\_and\\_university\\_rankings](https://en.wikipedia.org/wiki/College_and_university_rankings). [Diakses 18 December 2016].
- [8] Adina - Petruta Pavel, "Global University Rankings - A Comparative Analysis," *Procedia Economics and Finance* 26, pp. 54-63, 2015.
- [9] QS World University Rankings, "QS Global Academic Survey 2014," QS World University Rankings, 2014.
- [10] Vidya Rajiv Yeravdekar, Gauri Tiwari, "Global Rankings of Higher Education Institutions and India's Effective Non-Precense: Why Have World-Class Universities Eluded The Indian Higher Education System? And, How Worthwhile is The Indian Government's Captivation to Launch World Class Universities," *Procedia - Social and Behavioral Sciences* 157, pp. 63-83, 2014.

- [11] QS World University Rankings, "Annual QS Global Employer Survey," QS World University Rankings, 2014.
- [12] L. R. Raymond J. Mooney, "Content-based book recommending using learning for text categorization," dalam *DL '00 Proceedings of the fifth ACM conference on Digital libraries*, San Antonio, Texas, USA, 2000.
- [13] Wikipedia, "Wikipedia Scopus," 4 August 2016. [Online]. Available: <https://en.wikipedia.org/wiki/Scopus>. [Diakses 10 December 2016].
- [14] Elsevier, "Elsevier Developers," 1 January 2016. [Online]. Available: <http://dev.elsevier.com/>. [Diakses 10 December 2016].
- [15] Anne-Wil K. Harzing, Ron van der Wal, "Google Scholar as A New Source for Citation Analysis," *Ethics in Science and Environmental Politics*, pp. 61-73, 2008.
- [16] Wikipedia, "Wikipedia Google Scholar," 27 June 2017. [Online]. Available: [https://en.wikipedia.org/wiki/Google\\_Scholar](https://en.wikipedia.org/wiki/Google_Scholar). [Diakses 17 December 2016].
- [17] Wikipedia, "Wikipedia H-Index," 28 July 2017. [Online]. Available: <https://en.wikipedia.org/wiki/H-index>. [Diakses 28 July 2017].
- [18] Kementerian Riset, Teknologi dan Pendidikan Tinggi, "Profil, Fitur, Registrasi Sinta Pusat Index, Sitasi dan Kepakaran Terbesar di Indonesia," Kementerian Riset, Teknologi dan Pendidikan Tinggi, Jakarta, 2017.
- [19] Kementerian Riset, teknologi dan Pendidikan Tinggi, "Sinta Science and Technology Index," Kementerian Riset Teknologi dan Pendidikan Tinggi, 28 July 2017. [Online]. Available: <http://sinta1.ristekdikti.go.id/>. [Diakses 28 July 2017].
- [20] Kementerian Riset Teknologi dan Pendidikan Tinggi Republik Indonesia, "Kementerian Riset Teknologi dan Pendidikan Tinggi Republik Indonesia," 1 January 2016. [Online]. Available: <http://www.forlap.dikti.go.id/>. [Diakses 10 December 2016].
- [21] Arjun Thakur, A. L. Sangal, Harminder Bindra, "Quantitative Measurement and Comparison of Effects of Various SearchEngine Optimization Parameters on Alexa Traffic Rank," *International Journal of Computer Applications*, pp. 975-8887, 2011.
- [22] Wikipedia, "Wikipedia Python Programming Language," 28 July 2017. [Online]. Available:

- [https://en.wikipedia.org/wiki/Python\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)). [Diakses 28 July 2017].
- [23] Steven Bird, "NLTK: The Natural Language Toolkit," *Proceedings of the COLINGACL on Interactive presentation*, pp. 69-72, 2006.
  - [24] Tingting Wei, Yonghe Lu, Houyou Chang, Qiang Zhou, Xianyu Bao, "A Semantic Approach for Text Clustering using WordNet and Lexical Chains," *Expert Systems with Applications* 42, vol. 4, no. 1, pp. 2264-2275, 2015.
  - [25] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* 12, pp. 2825-2830, 2011.
  - [26] TextBlob, "TextBlob: Simplified Text Processing," 28 June 2017. [Online]. Available: <https://textblob.readthedocs.io/en/dev/>. [Diakses 28 June 2017].
  - [27] Wikipedia, "Wikipedia PHP," 28 July 2017. [Online]. Available: <https://en.wikipedia.org/wiki/PHP>. [Diakses 28 July 2017].
  - [28] Wikipedia, "Wikipedia MySQL," 28 July 2017. [Online]. Available: <https://en.wikipedia.org/wiki/MySQL>. [Diakses 28 July 2017].
  - [29] Daniel Jurafsky, James H. Martin, "Part-of-Speech Tagging," dalam *Speech and Language Processing*, 2016.
  - [30] Center of Teaching and Learning Stanford University, "Bloom's Taxonomy of Educational Objectives," Stanford University, California, 1981.
  - [33] William M. Darling, "A Theoretical and Practical Implementation Tutorial on Topic Modeling and Gibbs Sampling," 1 December 2011.
  - [34] Fony Revindasari, Riyanarto Sarno, Adhatus Solichah, "Traceability between business process and software component using Probabilistic Latent Semantic Analysis," 2016.
  - [35] David M. Blei, Jon D. McAuliffe, "Supervised Topic Model," *Advances in Neural Information Processing Systems*, 2007.
  - [37] Juan Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," 2003.
  - [38] Badan Akreditasi Nasional Perguruan Tinggi, "Pedoman Penyusunan Borang Akreditasi Institusi Perguruan Tinggi," Badan Akreditasi Nasional Perguruan Tinggi, Jakarta, 2011.

## LAMPIRAN

### Kode Sumber

```
1. //Document.java
2. package edu.smu.yl.plsa;
3.
4. import java.io.File;
5. import java.util.ArrayList;
6. import java.util.HashMap;
7. import java.util.List;
8. import java.util.Map;
9. import java.util.regex.Matcher;
10. import java.util.regex.Pattern;
11.
12. import edu.smu.yl.parser.StanfordTokenizer;
13. import edu.smu.yl.com.AlphanumFileComparator;
14. import edu.smu.yl.com.FileUtil;
15. import edu.smu.yl.com.Stopwords;
16. import java.util.Arrays;
17.
18.
```



```

19. public class Documents {
20.
21.     ArrayList<Document> docs;
22.     Map<String, Integer> termToIndexMap;
23.     ArrayList<String> indexToTermMap;
24.     Map<String,Integer> termCountMap;
25.
26.     public Documents(){
27.         docs = new ArrayList<Document>();
28.         termToIndexMap = new HashMap<String, Integer>();
29.
30.         indexToTermMap = new ArrayList<String>();
31.         termCountMap = new HashMap<String, Integer>();
32.     }
33.
34.     public void readDocs(String docsPath){
35.         File[] files = new File(docsPath).listFiles();
36.         System.out.print("");
37.         Arrays.sort(files, new AlphanumFileComparator());
38.
39.         for(File docFile : files){
40.             Document doc = new Document(docFile.getAbsolutePath(), termToIndexMap, indexToTermMap, termCountMap);
41.             docs.add(doc);
42.         }
43.     }
44.
45.     public static class Document {
46.
47.         private String docName;
48.         int[] docWords;
49.
50.         public Document(String docName, Map<String, Integer> termToIndexMap, ArrayList<String> indexToTermMap, Map<String, Integer> termCountMap){
51.             this.docName = docName;
52.             //Read file and initialize word index array
53.             ArrayList<String> docLines = new ArrayList<String>();
54.             ArrayList<String> words = new ArrayList<String>();
55.             List<String> tokenSents = new ArrayList<String>();
56.         }
57.     }
58. }

```

```

53. FileUtil.readLines(docName, docLines);
54. System.out.println(docName);
55. for(String line : docLines){
56. tokenSents = StanfordTokenizer.tokenizeSents(line
    );
57. for(String tokenSent : tokenSents){
58. FileUtil.tokenizeAndLowerCase(tokenSent, words);

59. }
60. }
61. //Remove stop words and noise words
62. for(int i = 0; i < words.size(); i++){
63. if(Stopwords.isStopword(words.get(i)) || isNoiseW
    ord(words.get(i))){
64. words.remove(i);
65. i--;
66. }
67. }
68. //Transfer word to index
69. this.docWords = new int[words.size()];
70. for(int i = 0; i < words.size(); i++){
71. String word = words.get(i);
72. if(!termToIndexMap.containsKey(word)){
73. int newIndex = termToIndexMap.size();
74. termToIndexMap.put(word, newIndex);
75. indexToTermMap.add(word);
76. termCountMap.put(word, new Integer(1));
77. docWords[i] = newIndex;
78. } else {
79. docWords[i] = termToIndexMap.get(word);
80. termCountMap.put(word, termCountMap.get(word) + 1
    );
81. }
82. }
83. words.clear();
84. }
85.
86. public boolean isNoiseWord(String string) {
87. // TODO Auto-generated method stub
88. string = string.toLowerCase().trim();
89. Pattern MY_PATTERN = Pattern.compile(".*[a-zA-
    Z]+.*");
90. Matcher m = MY_PATTERN.matcher(string);

```

```

91. // filter @xxx and URL
92. if(string.matches(".*www\\.\\.\\.") || string.matches
   (".*\\.\\.\\.com.*") ||
93. string.matches(".*http:.*") )
94. return true;
95. if (!m.matches()) {
96. return true;
97. } else
98. return false;
99. }
100.
101.     }
102.     }
103.     pLSA.java
104.     package edu.smu.yl.plsa;
105.
106.     import java.io.IOException;
107.     import java.util.ArrayList;
108.     import java.util.Collections;
109.     import java.util.Comparator;
110.     import java.util.List;
111.     import java.util.Random;
112.
113.     import edu.smu.yl.plsa.Documents.Document;
114.
115.     import edu.smu.yl.com.FileUtil;
116.     import edu.smu.yl.conf.ParamConfig;
117.     import edu.smu.yl.conf.PathConfig;
118.     import edu.smu.yl.database.DatabaseControl
119.     ler;
120.     import edu.smu.yl.plsa.CosineSimilarity;
121.     import java.io.BufferedWriter;
122.     import java.io.File;
123.     import java.io.FileWriter;
124.     import java.text.DecimalFormat;
125.     import org.apache.commons.lang3.ArrayUtils
126.     ;
127.
128.     /**
129.     * Class for PLSA model using EM
130.     *
131.     * @author yangliu

```

```

130.      * @blog http://blog.csdn.net/yangliuy
131.      * @mail yangliuyx@gmail.com
132.      */
133.      public class Plsa {
134.
135.          private int iters;
136.          private String universityName;
137.          private int saveStep;
138.
139.          private int beginSaveIters;
140.
141.          private int topicNum;
142.
143.          private int N; // number of docs
144.
145.          private int M; // number of terms
146.
147.          private int[][] docTermMatrix; // docT
ermMatrix
148.
149.          private double[][] docTopicPros;//p(z|
d)
150.
151.          private double[][] topicTermPros;//p(w
|z)
152.
153.          private double[][][] docTermTopicPros;
//p(z|d,w)
154.
155.
156.          public Plsa(String universityName) {
157.              //init parameters
158.              this.universityName = universityNa
me;
159.              topicNum = ParamConfig.topicNum;
160.              iters = ParamConfig.iteration;
161.              saveStep = ParamConfig.saveStep;
162.              beginSaveIters = ParamConfig.begin
SaveIters;
163.          }
164.
165.          public void initializeModel(Documents
docSet) {

```

```

166.         if (docSet == null) {
167.             throw new IllegalArgumentException(
168.                 tion("The documents set must be not null!"));
169.         }
170.         N = docSet.docs.size();
171.         M = docSet.indexToTermMap.size();

172.
173.         //element in docTermMatrix represe
174.         nts times the word appear in this document
175.         docTermMatrix = new int[N][M];
176.         //init docTermMatrix
177.         for (int docIndex = 0; docIndex <
178.             N; docIndex++) {
179.             Document doc = docSet.docs.get
180.             (docIndex);
181.             for (int word = 0; word < doc.
182.                 docWords.length; word++) {
183.                 int termIndex = docSet.doc
184.                 s.get(docIndex).docWords[word];
185.                 docTermMatrix[docIndex][te
186.                 rmIndex] += 1;
187.             }
188.         }

189.         docTopicPros = new double[N+1][top
190.             icNum + 2];
191.         topicTermPros = new double[topicNu
192.             m][M];
193.         docTermTopicPros = new double[N][M
194.             ][topicNum];

195.         //init p(z|d),for each document th
196.         e constraint is sum(p(z|d))=1.0
197.
198.         for (int i = 0; i <= N; i++) {
199.             double[] pros = randomProbilit
200.             ies(topicNum);
201.             for (int j = 0; j < topicNum;
202.                 j++) {
203.                 docTopicPros[i][j] = pros[
204.                     j];

```

```

194.         }
195.
196.     }
197.     //init p(w|z),for each topic the c
    onstraint is sum(p(w|z))=1.0
198.     for (int i = 0; i < topicNum; i++)
199.     {
200.         double[] pros = randomProbit
ies(M);
201.         for (int j = 0; j < M; j++) {
202.             topicTermPros[i][j] = pros
[j];
203.         }
204.     }
205.
206.     /**
207.      *
208.      * Plsa inference using EM
209.      *
210.      * @param docs all documents
211.      * @throws IOException
212.      */
213.     public void inferenceModel(Documents d
ocSet) throws IOException {
214.         //use em to estimate params
215.         //save model according to model pa
rameters
216.
217.         if (iters < saveStep + beginSaveIt
ers) {
218.             System.err.println("Error: the
number of iterations should be larger than "
219.                 + (saveStep +
beginSaveIters));
220.             System.exit(0);
221.         }
222.         for (int i = 0; i <= iters; i++) {
223.             System.out.println("Iteration
" + i + " ----- ");
224.             if ((i >= beginSaveIters)

```

```

225.         && (((i - beginSaveIter
rs) % saveStep) == 0)) {
226.             // Saving the model
227.             System.out.println("Saving
model at iteration " + i + " ... ");
228.
229.
230.
231.             GenerateMa
x();
232.             System.out
.println(+ i);
233.             GenerateBo
bot(i);
234.             saveIteratedModel(i, docSe
t);
235.         }
236.         em();
237.         System.out.println("After E st
ep and M step, the new log likelihood is " + comp
uteLogLikelihood(docSet));
238.     }
239.
240.     System.out.println("done");
241. }
242.
243. private void GenerateMax(){
244.     for(int i=0;i<N;i++){
245.         double tempMax = docTopicPros[
i][0];
246.         for(int j=1;j<topicNum;j++){
247.             if(docTopicPros[i][j] > t
empMax )
248.                 tempMax = docTopicPros
[i][j];
249.         }
250.         docTopicPros[i][topicNum] = te
mpMax;
251.     }
252. }
253. private void GenerateBobot(int iterati
on){
254.     double bobot , sum = 0;

```

```

255.         for(int i=0;i<N;i++){
256.             for(int j=0;j<topicNum;j++){
257.                 if(docTopicPros[i][j] ==
docTopicPros[i][topicNum] ){
258.                     bobot = docTopicPros[i
][j];
259.                     bobot *= ((j+1) * 10);

260.
261.                     docTopicPros[i][topicN
um + 1] = bobot;
262.                     sum += bobot;
263.                     break;
264.                 }
265.             }
266.         }
267.         sum /= N;
268.         for(int i=0;i<=topicNum;i++)
269.             docTopicPros[N][i] = 0;
270.         docTopicPros[N][topicNum +1] = sum
;
271.         if(iteration/10 == N-1)
272.             new DatabaseController().Inser
t(sum , universityName);
273.         }
274.
275.         private double computeLogLikelihood(Doc
uments docSet) {
276.             // TODO Auto-
generated method stub
277.             /*
278.              * Compute the log likelihood of g
eneration the corpus
279.              *
280.              *  $L = \sum_i \{n(d_i) * [\sum_j (n(d_i, w_j) / n(d_i) * \log \sum_k p(z_k|d_i) * p(w_j|z_k))]\}$ 
281.              *
282.              */
283.             double L = 0.0;
284.             for (int i = 0; i < N; i++) {
285.                 double docISize = docSet.docs
.get(i).docWords.length;

```



```

286.         double sumM = 0.0;
287.         for (int j = 0; j < M; j++) {
288.             double sumK = 0.0;
289.             for(int k = 0; k < topicN
um; k++){
290.                 sumK += docTopicPros[
i][k] * topicTermPros[k][j];
291.             }
292.             //System.out.println("sum
K: " + sumK);
293.             sumM += (double)docTermMa
trix[i][j] / docISize * Math.log10(sumK);
294.         }
295.         //System.out.println("sumM: "
+ sumM);
296.         L += docISize * sumM;
297.     }
298.     return L;
299. }
300.
301. /**
302.  *
303.  * EM algorithm
304.  *
305.  */
306. private void em() {
307.     /*
308.      * E-
step, calculate posterior probability  $p(z|d, w, \&)$ , &
is
309.      * model params( $p(z|d)$ ,  $p(w|z)$ )
310.      *
311.      *  $p(z|d, w, \&) = p(z|d) * p(w|z) / \text{sum}(p(
z'|d) * p(w|z'))$ 
312.      *  $z'$  represent all possible topic
313.      *
314.      */
315.     for (int docIndex = 0; docIndex <
N; docIndex++) {
316.         for (int wordIndex = 0; wordIn
dex < M; wordIndex++) {
317.             double total = 0.0;

```

```

318.         double[] perTopicPro = new
        double[topicNum];
319.         for (int topicIndex = 0; t
opicIndex < topicNum; topicIndex++) {
320.             double numerator = doc
TopicPros[docIndex][topicIndex]
321.                 * topicTermPro
s[topicIndex][wordIndex];
322.             total += numerator;
323.             perTopicPro[topicIndex
] = numerator;
324.         }
325.
326.         if (total == 0.0) {
327.             total = avoidZero(tota
1);
328.         }
329.
330.         for (int topicIndex = 0; t
opicIndex < topicNum; topicIndex++) {
331.             docTermTopicPros[docIn
dex][wordIndex][topicIndex] = perTopicPro[topicIn
dex]
332.                 / total;
333.         }
334.     }
335. }
336.
337. //M-step
338. /*
339.      * update  $p(w|z), p(w|z) = \frac{\sum(n(d', w) * p(z|d', w, \&))}{\sum(\sum(n(d', w') * p(z|d', w', \&)))}$ 
340.      *
341.      * d' represent all documents
342.      * w' represent all vocabularies
343.      *
344.      *
345.      */
346.     for (int topicIndex = 0; topicInde
x < topicNum; topicIndex++) {
347.         double totalDenominator = 0.0;

```

```

348.         for (int wordIndex = 0; wordIndex < M; wordIndex++) {
349.             double numerator = 0.0;
350.             for (int docIndex = 0; docIndex < N; docIndex++) {
351.                 numerator += docTermMatrix[docIndex][wordIndex]
352.                     * docTermTopicPros[docIndex][wordIndex][topicIndex];
353.             }
354.
355.             topicTermPros[topicIndex][wordIndex] = numerator;
356.
357.             totalDenominator += numerator;
358.         }
359.
360.         if (totalDenominator == 0.0) {
361.             totalDenominator = avoidZero(totalDenominator);
362.         }
363.
364.         for (int wordIndex = 0; wordIndex < M; wordIndex++) {
365.             topicTermPros[topicIndex][wordIndex] = topicTermPros[topicIndex][wordIndex]
366.                 / totalDenominator;
367.         }
368.     }
369.     /*
370.     * update p(z|d), p(z|d)=sum(n(d,w')*p(z|d,w'&))/sum(sum(n(d,w')*p(z'|d,w',&)))
371.     *
372.     * w' represent all vocabularies
373.     * z' represent all topics
374.     *
375.     */
376.     for (int docIndex = 0; docIndex < N; docIndex++) {

```

```

377.                                     //actually equal sum(w) of thi
    s doc
378.                                     double totalDenominator = 0.0;

379.                                     for (int topicIndex = 0; topic
Index < topicNum; topicIndex++) {
380.                                         double numerator = 0.0;
381.                                         for (int wordIndex = 0; wo
rdIndex < M; wordIndex++) {
382.                                             numerator += docTermMa
trix[docIndex][wordIndex]
383.                                                 * docTermTopic
Pros[docIndex][wordIndex][topicIndex];
384.                                         }
385.                                         docTopicPros[docIndex][top
icIndex] = numerator;
386.                                         totalDenominator += numera
tor;
387.                                     }
388.
389.                                     if (totalDenominator == 0.0) {
390.                                         totalDenominator = avoidZe
ro(totalDenominator);
391.                                     }
392.
393.                                     for (int topicIndex = 0; topic
Index < topicNum; topicIndex++) {
394.                                         docTopicPros[docIndex][top
icIndex] = docTopicPros[docIndex][topicIndex]
395.                                             / totalDenominator
;
396.                                     }
397.                                 }
398.        }
399.
400.        /**
401.         *
402.         *
403.         * Get a normalize array
404.         *
405.         * @param size
406.         * @return

```

```

407.          */
408.      public double[] randomProbabilities(int
      size) {
409.          if (size < 1) {
410.              throw new IllegalArgumentException(
      tion("The size param must be greate than zero"));
411.          }
412.          double[] pros = new double[size];
413.
414.          int total = 0;
415.          Random r = new Random(2);
416.          for (int i = 0; i < pros.length; i
      ++){
417.              //avoid zero
418.              pros[i] = r.nextInt(size) + 1;
419.
420.              total += pros[i];
421.          }
422.
423.          //normalize
424.          for (int i = 0; i < pros.length; i
      ++){
425.              pros[i] = pros[i] / total;
426.          }
427.
428.          return pros;
429.      }
430.
431.      /**
432.       *
433.       * @return
434.       */
435.      public double[][] getDocTopics() {
436.          return docTopicPros;
437.      }
438.
439.      /**
440.       *
441.       * @return
442.       */

```

```

443.         public double[][] getTopicWordPros() {
444.             return topicTermPros;
445.         }
446.
447.         /**
448.          *
449.          * Get topic number
450.          *
451.          *
452.          * @return
453.          */
454.         public Integer getTopicNum() {
455.             return topicNum;
456.         }
457.
458.         /**
459.          *
460.          * avoid zero number.if input number i
461.          * s zero, we will return a magic
462.          * number.
463.          *
464.          */
465.         private final static double MAGICNUM =
466.             0.0000000000000001;
467.         public double avoidZero(double num) {
468.             if (num == 0.0) {
469.                 return MAGICNUM;
470.             }
471.
472.             return num;
473.         }
474.
475.
476.
477.         pLSAtest.java
478.         package edu.smu.yl.plsa;
479.
480.         import java.io.IOException;
481.

```

```

482.     import edu.smu.yl.conf.ParamConfig;
483.     import edu.smu.yl.conf.PathConfig;
484.     import java.io.File;
485.     import java.io.FilenameFilter;
486.     import java.util.logging.Level;
487.     import java.util.logging.Logger;
488.
489.     /**
490.      * Test Class for PLSA model using EM algo
491.      *
492.      * @author yangliu
493.      * @blog http://blog.csdn.net/yangliuy
494.      * @mail yangliuyx@gmail.com
495.      */
496.
497.     public class PlsaTest {
498.
499.         public static void main(String[] args)
500.             throws IOException {
501.             String path = PathConfig.oriDataPa
502.             th;
503.             File file = new File(path);
504.             String[] directories = file.list(n
505.             ew FilenameFilter() {
506.                 @Override
507.                 public boolean accept(File curre
508.                 nt, String name) {
509.                     Documents docSet = new Documents()
510.                     ;
511.                     System.out.println("0 Read Docs ..
512.                     .");
513.                     docSet.readDocs(PathConfig.oriData
514.                     Path + name);
515.                     System.out.println("docSet: " + do
516.                     cSet.docs.size());
517.                     Plsa model = new Plsa(name);
518.                     System.out.println("1 Initialize t
519.                     he model ...");
520.                     model.initializeModel(docSet);
521.                     System.out.println("2 Learning and
522.                     Saving the model ...");
523.                     try {

```

```

514.                model.inferenceModel(doc
Set);
515.                } catch (IOException ex) {
516.                    Logger.getLogger(PlsaTes
t.class.getName()).log(Level.SEVERE, null, ex);
517.                }
518.                System.out.println("3 Output the f
inal model ...");
519.                try {
520.                    model.saveIteratedModel(
ParamConfig.iteration, docSet);
521.                } catch (IOException ex) {
522.                    Logger.getLogger(PlsaTes
t.class.getName()).log(Level.SEVERE, null, ex);
523.                }
524.                System.out.println("Done!");
525.                return new File(current, name)
.isDirectory();
526.            }
527.        });
528.
529.
530.
531.    }
532. }

```



*[Halaman ini sengaja dikosongkan]*



## BIODATA PENULIS

Donny Aliyanto atau yang biasa disapa dengan nama Doni, lahir di Surabaya pada tanggal 6 Mei 1995. Ketua dari Komunitas Blogger Bangkalan yang berdiri pada tahun 2016 dan saat ini sedang menempuh pendidikan Sarjana di Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya angkatan tahun 2013.

Memiliki ketertarikan di bidang Internet Marketing yang merupakan salah cara mengubah hobby menjadi dollar. Internet Marketing adalah sebuah seni pemasaran secara online bisa berupa produk, jasa, atau pengiklanan. Jika anda memiliki hobby yang sama dengan saya dan ingin mengubah hobby anda menjadi pekerjaan anda, silahkan kontak saya : [doniemax@yahoo.com](mailto:doniemax@yahoo.com) atau di [donnyaliyanto.web.id](http://donnyaliyanto.web.id)